Vision Enhancement for Autonomous Vehicles: A Deep Learning approach for Anticipating and Adapting to Adverse Weather Conditions

by

SADMAN TAHMID CHOWDHURY
24341124
ARIFA HUMAIRA
21341017
ILHAN BIN AHSAN
21201514
MASRURAH MEHRNAZ
21201771

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
December, 2025.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

SADMAN TAHMID CHOWDHURY 24341124	**ARIFA HUMAIRA** 21341017
ILHAN BIN AHSAN 21201514	**MASRURAH MEHRNAZ** 21201771

Approval

The thesis titled "An Efficient Deep Learning Approach Leveraging Dual Task Evaluation For Vision Enhancement Of Autonomous Vehichles in Adverse Weather Conditions" submitted by

1. Sadman Tahmid Chowdhury (24341124)
2. Arifa Humaira (21341017)
3. Ilhan Bin Ahsan (21201514)
4. Masrurah Mehrnaz (21201771)

Of Summer, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on October 10, 2025.

Examining Committee:
Supervisor:
(Member)

Dr. Md. Ashraful Alam
Associate Professor
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Dr. Md Golam Robiul Alam
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Associate Professor
Department of Computer Science and Engineering
Brac University

Ethics Statement

This research was conducted in compliance with ethical standards and with the basic tenets of academic integrity. All processes involved in collecting, analyzing, and reporting information were conducted with integrity and transparency, and in compliance with ethical standards.

The research did not include any human subjects, and no individual and sensitive information was included in it. All secondary sources utilized have been duly referenced, with proper credit taken for the original sources and authors.

Moreover, this study conforms to both academic and institution-established ethical standards. Any and all conflicts of interest have been disclosed in full, and anti-plagiarism protocols have been taken with utmost care through proper use of references.

The thesis is prepared with a genuine desire to contribute to a specific field of study, and in doing so, with integrity in terms of intellectual property, privacy, and ethics in research.

Abstract

The camera-based perception reliability in autopilot vehicles in poor weather conditions like when it is pouring, there is thick fog, or snow is severely affected. These visual illusions cause missed detections, misclassifications and latent responses hence compromising safety. The current perception systems are usually effective in clear weather but their performance declines drastically during extreme or transitional weather and this poses an operational risk. This thesis presents a framework, which examines the effects of bad weather on visual perception through two complementary tasks, which are weather classification and object detection. An EfficientNet-B3 model with multi-class weather classification was fine-tuned and a pretrained YOLOv8m (COCO weights) tested the strength of object detection under the same weather conditions. This framework measures the impact of environmental degradation on the reliability of perception by correlating classification outputs and detection performance. Moreover, a Custom EfficientNet-B3 classifier was created and had a two-layered classification head ($1536 \rightarrow 256 \rightarrow 4$) with droupout. To balance between feature preservation and domain adaptation, the model uses progressive backbone unfreezing in two training phases. Such a setup was more accurate, and it surpassed the baseline in terms of improved feature extraction and regularization and was also computationally affordable to run on an embedded platform. Our custom driving scenes in Bangladesh with augmented physics based depth-aware fog rendering and CycleGAN generated rain and snow condition form the dataset used in this study. Moreover, for greater detection robustness in adverse weather conditions, the YOLOv8m model was improved with the addition of a Convolutional Block Attention Module (CBAM) subsequent to the last block, C2f, within the model's backbone. CBAM facilitates the adaptive recalibration of channel-wise as well as spatial attributes, allowing the network to pay attention to appropriate object attributes despite adverse weather conditions. The overall objective is to provide a weather-aware evaluation methodology that enhances understanding of autonomous driving reliability under diverse and challenging environments.

Keywords: adverse weather, object detection, deep learning, EfficientNet-B3, YOLOv8m.

Dedication

This thesis is devoted to the students and the faculty members of Department of Computer Science and Engineering at Brac University whose determination to excel and explore academically has given us the basis to conduct this research. The faculty and staff guidance and support has played a great role in shaping our knowledge and skills to accept challenges in the discipline of computer science. It is also dedicated to the world computer vision community and enthusiasts who have tirelessly worked on improvements in artificial intelligence and image processing that has inspired and challenged the technology boundaries.The work you have done, the research you have conducted, and your transparent cooperation with us has given the foundation at which we develop in our current state. Hopefully, this contribution will be a little step in the joint process of knowledge and innovation.

Acknowledgement

We would like to say the greatest thanks to all the people who have been giving us the support and advice during this research work. We would like to thank first of all our thesis supervisor, Dr. Md. Ashraful Alam, and his precious guidance, compassion and encouragement. His critical remarks and unyielding advocacy also played the most beneficial role in the development of this research and its successful completion. In this connection, we wish to acknowledge the fact that our thesis supervising panel has helped us to improve our work by critically assessing it and providing professional advice and unwavering support. Their determination to achieve academic excellence has immensely improved the quality of our work. We are also grateful to our fellow thesis colleagues, whose cooperation in general, interesting discussions, and dedication to this project was significant. Their support and coordination have made this a rewarding and enriching experience.

In addition, we thankfully acknowledge the support of the required academic resources, research facilities, and study-conducive environment by the Department of Computer Science and Engineering of Brac University. The exposure and knowledge received from this department have helped greatly towards our academic and professional development.

Finally, we would like to thank our families and friends for their constant encouragement, patience, and moral support during this project. Their confidence in our capabilities has always been a great motivating force.

This thesis is the result of the cumulative inputs, advice, and support of all these individuals, to whom we are greatly indebted.

Table of Contents

Declaration **i**

Approval **ii**

Ethics Statement **ii**

Abstract **iv**

Dedication **v**

Acknowledgment **vi**

Table of Contents **vii**

List of Tables **ix**

List of Figures **x**

1 Introduction **1**
- 1.1 Background . . . 1
- 1.2 Rationale of the study and motivation . . . 2
- 1.3 Problem Statement . . . 2
- 1.4 Research Question . . . 3
- 1.5 Impact and Significance . . . 3
- 1.6 Research Objectives . . . 3
- 1.7 Overview of the Methodology . . . 4
- 1.8 Architectural and Methodological Challenges . . . 4

2 Literature Review **6**
- 2.1 Preliminaries . . . 6
- 2.2 Review of Existing Research . . . 6
  - 2.2.1 Weather Classification . . . 6
  - 2.2.2 Object Detection . . . 8
  - 2.2.3 Object Detection in Adverse Weather . . . 9
  - 2.2.4 Attention Mechanism in YOLOv8 for Detection . . . 9
  - 2.2.5 Bottleneck Classifiers . . . 11
  - 2.2.6 Synthetic Data Generation . . . 11
  - 2.2.7 Transfer Learning . . . 12
  - 2.2.8 Domain Adaptation . . . 12
  - 2.2.9 Uncertainty Quantification . . . 13

2.2.10 South Asian Deployment Context 13
2.2.11 Research Gaps 14
2.3 Emerging Design Patterns and Thesis positioning. 14

3 Requirements, Impacts and Constraints 16
3.1 Final Specifications and Requirements 16
3.2 Societal Impact 16
3.3 Environmental Impacts 17
3.4 Ethical Issues 17
3.5 Risk Management 18
3.6 Economic Analysis 18

4 Methodology 20
4.1 Overview of Proposed Approach 20
4.2 Dataset Development 21
4.2.1 Base Dataset Collection 21
4.2.2 Data Organization and Splitting 22
4.2.3 Weather Augmentation Pipeline 22
4.2.4 Final Dataset Composition 25
4.3 Weather Classification Models 25
4.3.1 EfficientNet-B3 Base Model 25
4.3.2 Custom EfficientNet-B3 (Two-layer classifier design Variant) . 26
4.4 Object Detection Framework 28
4.4.1 Baseline YOLOv8m Configuration 29
4.4.2 CBAM-Enhanced YOLOv8m Architecture 30
4.4.3 Evaluation Metrics and Protocol 32

5 Results and Discussion 34
5.1 Dataset Analysis 34
5.2 Weather Classification Results 35
5.2.1 Evaluation Metrics 35
5.2.2 EfficientNet-B3 Baseline Results 36
5.2.3 Custom EfficientNet-B3 Results 38
5.3 Object Detection Results 40
5.3.1 Aggregate Results 40
5.3.2 Weather-Stratified Detection Results 42
5.4 Discussion 46
5.4.1 High and Balanced Classification Accuracy 47
5.4.2 Predictable Detection Degradation 47
5.4.3 Summary of Key Insights 48

6 Conclusion 49
6.1 Limitations 49
6.2 Future Works 49
6.3 Overall Conclusion 50

References 51

List of Tables

4.1 Comparison between Baseline and Custom EfficientNetB3 Models . . 28
4.2 Hyperparameters for Two-Stage Training 28

5.1 Aggregate detection metrics across all weather conditions using baseline YOLOv8m. 41
5.2 Aggregate detection metrics across all weather conditions using CBAM-enhanced YOLOv8m. 41
5.3 Detection performance under clear weather conditions. 42
5.4 Detection performance under foggy conditions. 43
5.5 Detection performance under rainy conditions. 43
5.6 Detection performance under snowy conditions. 44
5.7 Mean mAP@0.5 comparison across all weather conditions. 45

List of Figures

4.1 Diagram showing the specifications of the Baseline models and the Custom model 27

5.1 Sample Image from our dataset. The clear image is the ground truth and the rest are augmented 34
5.2 Sample Image from our dataset showing manually annotated bounding boxes using Roboflow software 35
5.3 Confusion matrix for EfficientNet-B3 baseline on the test set. 37
5.4 Training and validation loss curves for EfficientNet-B3. 37
5.5 Training and Validation accuracy curves for EfficientNet-B3. 38
5.6 Confusion matrix for Customized Model on the test set. 39
5.7 Training and validation loss curves for Custom Model. 39
5.8 Training and Validation accuracy curves for Custom Model 40
5.9 Mean IoU distribution across weather conditions (Baseline YOLOv8m). 45
5.10 Mean IoU distribution across weather conditions (Customized YOLOv8m). 46

Chapter 1

Introduction

1.1 Background

On January 15, 2023, heavy fog outside San Francisco contributed to a self-driving Tesla colliding with a stationary emergency vehicle. The vehicle's perception system failed to detect the obstacle early enough to prevent the crash. These examples indicate the systematic failure of visionbased autonomy: models of perception that are highly accurate, above 95%, in clear environments can fall below 65% when disparities are decreased by fog and object boundaries are obscured[10], [18]. The consequences include vehicle misclassification, missed pedestrians, and delayed emergency braking each of which is extremely dangerous to safety. To make autonomous vehicles practical in practical use, the perception systems should not malfunction due to fluctuating weather conditions. Camera based vision is cheap and information intensive but very sensitive to fog, rain and snow. Such weather phenomena cause poor image quality in terms of contrast, blurred edges, and distorted color, leading to the high perception latency and error [7]. This is made more difficult by small datasets that reflect dissimilar geographical and environmental settings. Commonly available publicly available data, including BDD100K and ACDC, mostly represent Western traffic and infrastructure, which only gives a small amount of information about the South Asian urban setting like Dhaka, Bangladesh.
The recent progress in deep learning provides indicators to suggest that the separate assessment of many perception tasks and the following correlation of their outputs can be used as a way of getting more insight into the manner in which environmental conditions influence the reliability of autonomous driving. The joint analysis of weather classification and object detection on the same data allows quantifying the effect of various weather regimes on the reduction of detection performance. EfficientNet architectures are known for strong accuracy–efficiency trade-offs in image classification, while YOLO models provide real-time object detection suitable for embedded use. However, the impact of these models under adverse weather specific to Bangladesh remains underexplored.
This work fills this gap with the proposal of a **weather-aware object detection system** that improves environmental classification and detection accuracy in adverse weather. Specifically, this system utilizes a personalized **EfficientNet-B3** neural network model for environmental classification, as well as a customized **YOLOv8m** model with attention mechanism for object detection. In the YOLOv8m, the addition of a CBAM attention mechanism enhances its capabilities with the in-

corporation of a CBAM module right after its last C2f block, allowing for channel as well as spatial attention that promotes adaptability in feature recalibration, thus coping with environmental occlusions affecting visibility in adverse weather. This project utilizes a generated dataset composed of Bangladeshi driving environmental scenes with weather effects generated through physics-based depth-inclusive fog simulation, as well as CycleGAN-produced rain and snow conditions.

1.2 Rationale of the study and motivation

Perception systems that achieve over 95% detection accuracy in clear weather often drop below 65% under dense fog [10], [18]. Fog scatters light through Mie scattering, suppressing contrast by 40–60% and shifting color temperature toward a blue-gray spectrum. These effects destabilize convolutional edge detectors and reduce gradient energy at object boundaries. The result: pedestrians vanish from bounding boxes, vehicles are misclassified, and emergency braking is triggered too late.
For autonomous driving to be viable in all conditions, perception systems must remain robust under such weather-induced degradations. Conventional approaches, however, rarely account for these scattering effects in model design, leaving safety-critical systems vulnerable precisely when visibility is most limited.

1.3 Problem Statement

Regional Bias

Benchmark datasets include BDD100K and ACDC which mainly represent Western driving environments BDD100K, ACDC. As an example, BDD100K has over 70 percent highway and suburban images and less than 10% represent heavy urban traffic. This is the case with Dhaka, where around 65% of road life is concentrated in urban situations of heterogeneous actors who have no road signage, including rickshaws, CNG auto-rickshaws, and buses, as well as pedestrians. Models trained on Western highways have a hard time generalizing to these conditions, which have significantly different patterns of occlusion and road dynamics.

Natural Fog Capture Constraints

To gather natural mist data, multi-season driving is important to ensure that data on variable density, visibility ranges and day-time are obtained. These are resource-consuming and unsafe efforts, as the impoverished states decrease the human fallback ability. Therefore, not many big datasets of fogs are present. Synthetic augmentation is scalable, but GAN-based approaches introduce trade-offs: they yield unnatural edge halos and inconsistent depth cues, whereas physics-based rendering needs to use correct depth maps and calibration to prevent over-smoothing or too-large contrast[18].

Gap Statement

At present, no large-scale, fog-augmented dataset exists for South Asian urban driving. This absence forces perception models to extrapolate from Western benchmarks to Dhaka's traffic, which exhibits inverse scene distributions and unique occlusion challenges. The lack of regionally relevant, weather-augmented datasets is therefore a central barrier to developing reliable fog perception systems for Bangladesh.

1.4 Research Question

These dataset and architectural challenges motivate the following research question: *How does a weather-aware object detection framework integrating EfficientNet-B3 weather classification with CBAM-enhanced YOLOv8m object detection perform in terms of detection robustness, weather-specific performance metrics, and risk stratification under adverse weather conditions in Bangladesh urban environments?*
Three sub-questions follow:

1. Does the CBAM-enhanced YOLOv8m architecture demonstrate improved detection metrics across adverse weather conditions compared to baseline YOLOv8m, and can weather classification outputs guide performance analysis?

2. Can a Bangladesh-specific dataset with synthetically generated adverse weather reduce the performance gap between clear-weather and adverse-weather detection relative to models trained on Western benchmarks?

3. What is the computational trade-off between the baseline EfficientNet-B3 classifier and the custom two-layer variant in terms of parameter count, training time, and classification accuracy?

1.5 Impact and Significance

Reliable perception addresses a fundamental safety requirement for autonomous deployment. Regionally, this research establishes a Bangladesh-specific benchmark that broadens the scope of autonomous vehicle perception beyond Western contexts. Technically, it demonstrates how attention mechanisms integrated into YOLOv8m's backbone can improve detection robustness under adverse weather, offering insights into architectural enhancements for weather-aware perception in resource-constrained autonomous systems.

1.6 Research Objectives

The objective of this study is to design and evaluate a dual-task **evaluation framework** capable of assessing perception robustness under adverse weather using real driving scenarios from Bangladesh. Specifically, the objectives are as follows:

1. **Address data scarcity:** Develop a Bangladesh-based dataset for adverse weather, using physics-based depth-aware fog synthesis and CycleGAN-based rain and snow generation to enhance local driving scenes.

2. **Develop the weather-aware detection framework:** Train a customized EfficientNet-B3 for weather classification and enhance YOLOv8m by integrating CBAM after the final C2f block in the backbone, then fine-tune this modified architecture for robust object detection across weather conditions.

3. **Establish baselines:**Train and evaluate standard EfficientNet-B3 for weather classification and baseline YOLOv8m for object detection independently, then compare against CBAM-enhanced YOLOv8m to quantify performance improvements.

4. **Compare performance:** Analyze how detection metrics (mAP, precision, recall) vary across weather classes for both baseline and CBAM-enhanced YOLOv8m, and compare baseline results against our custom weather classification model variant to demonstrate trade-offs.

5. **Discuss improvements:** Identify limitations and propose future research directions, including improved data annotation, active learning for detection labels, and the potential integration of weather-aware architectures.

1.7 Overview of the Methodology

The intended methodology creates a weather-aware detection framework to quantify to what extent adverse weather affects visual perception in Bangladesh city driving conditions. It combines two components: a weather categorization model (EfficientNet-B3 and a custom two-layer variation) trained on augmented sets of clear, fog, rain, and snow conditions, and a CBAM-enhanced YOLOv8m detector that integrates attention mechanisms after the final C2f block in the backbone, initialized with COCO pre-trained weights and fine-tuned for robust detection across weather conditions. The classifier is trained with a two-stage curriculum with selective unfreezing and aggressive dropout to optimize accuracy and efficiency, while the modified YOLOv8m detector is fine-tuned to improve detection performance under adverse weather through attention-based feature enhancement. A synthetic dataset (15,940 images) is created with physics-based fog simulation and GAN-based rain/snow augmentation. Performance analyses comparing baseline and CBAM-enhanced YOLOv8m across weather classes measure detection metrics (mAP, precision, recall) to quantify visibility-induced degradation and attention mechanism effectiveness, providing interpretable and deployment-ready insights into weather-aware perception performance.

1.8 Architectural and Methodological Challenges

Single-task Limitations

Most detection pipelines lack architectural mechanisms to adaptively respond to adverse weather conditions. Common detectors use fixed feature extraction and thresholds that are set to operate with clear weather conditions, leading to missed detections of fog covered pedestrians or rain damaged road signs. One way out of this is attention mechanisms such as CBAM which are able to recalibrate adaptive features so that models can pay attention to useful visual data in case the environment worsens image quality. Nevertheless, their application to Bangladesh-specific adverse weather conditions has not been empirically investigated to fit in the YOLOv8 architecture.

Fog-specific Degradation

The effects of scattering that have been mentioned above make the task of training robust models more challenging. Poor contrast, poor edges, and depth uncertainty decrease the confidence of the detector and sensor noise becomes more difficult to tell whether it is a real fog artifact. To achieve real-time inference (?30 FPS) under these conditions, architectures that trade-off representational capacity against computational efficiency are necessary, which limits the available choices of architecture to be used on embedded systems.

Chapter 2

Literature Review

2.1 Preliminaries

Autonomous perception systems use convolutional neural networks (CNNs) to extract hierarchical features from visual inputs for the purpose of environmental classification and object detection[9], [13]. EfficientNet uses compound scaling to scale network depth, width, and input resolution simultaneously through a unified coefficient ϕ and maintains optimal trade-offs between representational capacity and computational cost through mobile inverted bottleneck convolutions (MBConv) with squeeze-and-excitation (SE) blocks[14]. Single-stage detectors such as YOLO estimate bounding boxes and class probabilities directly from dense feature maps; YOLOv8 extends this paradigm through anchor-free formulation—predicting object centers and dimensions without the need for predefined anchor grids—along with decoupled detection heads that split classification and localization streams, with Task-Aligned Learning (TAL) dynamically designating the positive samples by both classification confidence and localization quality[12], [15]. Bad weather causes covariate shift between training and test distributions: rain creates streaks and specularities, fog weakens contrast and suppresses high spatial frequencies, and low light lowers signal-to-noise ratio. Transfer learning counters this through the pretrained learning on large source datasets and the fine-tuning on the target-domain samples with domain-aligned data augmentation, whereas unpaired image-to-image translation through the use of CycleGAN facilitates style transfer between the weather domains through the use of cycle-consistency and the adversarial losses[43]. Quantification of uncertainty—separating aleatoric (the data noise) from epistemic (the model uncertainty)—facilitates risk-aware decision-making essential for the safety-critical autonomous systems running under variable environmental conditions, providing the confidence estimates that control enhancement strength and communicate the perception reliability information to the planning modules.

2.2 Review of Existing Research

2.2.1 Weather Classification

Random changes in the environment alter image formation and the distribution of features learned by perception backbones. Rain produces temporally persistent streaks and more specular road surfaces; fog veils luminance and suppresses high spatial frequencies; low light reduces dynamic range and amplifies sensor noise; and glare saturates regions and induces blooming. These effects lower gradient energy at object edges, shift color statistics, and destabilize intermediate feature maps, degrading recall and localization in recognition and segmentation modules. Surveys consistently frame adverse conditions as a first-order failure mode rather than a corner case, with issues propagating from optics to annotation practice and evaluation protocol [5]. Field deployments indicate that camera-only and fused pipelines are both sensitive to extremes of illumination and precipitation despite maturity of sensors [7]. Restoration-oriented reviews further support the idea that condition awareness is most effective when used *upstream*, so downstream models are trained and deployed on inputs whose statistics match training-time distributions [9], [12], [15], [43].

Environmental classification serves as an upstream gateway for fast, conditional routing and parameterization. efficientnet-b3 is attractive here because compound scaling preserves a strong accuracy–latency trade-off suitable for embedded use: depth, width, and resolution are scaled in a coordinated fashion via a single coefficient ϕ and (α, β, γ), respecting a target compute budget $\alpha{\cdot}\beta^2{\cdot}\gamma^2$. This synchronized scaling avoids representational bottlenecks from single-dimension scaling and yields discriminative power proportional to input resolution. Deployment-oriented studies on real-time weather/illumination recognition emphasize lightweight backbones to keep condition prediction secondary to detection [25]. Two design rules emerge. One path uses explicit routing: the classifier's output triggers deraining/dehazing/low-light enhancement and tunes detector thresholds/NMS for the predicted regime, akin to weather-aware mixtures of experts [16]. The other path encodes condition awareness via a regime-sensitive unifier that maps diverse inputs to a normalized representation for a single fixed detector [43]. Both improve robustness over condition-agnostic baselines, provided condition inference is stable across regime transitions and severity shifts.

Backbone comparisons increasingly pit efficientnet-b3 against alternatives. At matched input resolutions, B4 typically offers higher accuracy with lower latency than ResNet-50/101, and better calibration under rapid contrast/illumination changes than ultra-light MobileNetV2/V3 [23], [25]. MBConv+SE yields channel-wise attention and depthwise separability, improving feature efficiency per FLOP in low-texture weather; additional depth/width headroom helps separate adjacent regimes (e.g., haze vs. light fog) without exceeding on-board budgets [25]. Studies on routing find diminishing returns from very heavy backbones once routing latency is accounted for, while very small mobile backbones can underfit severity gradations that matter for enhancement strength [19], [25].

Domain specificity is essential for external validity. South Asian conditions—heavy monsoon rain, dense winter haze, rainy nights, low-contrast dusk streets, and diverse roadside artifacts—are poorly represented in common public benchmarks. Local theses capture day/night, wet/dry surfaces, urban/rural roads, and region-specific obstacles to better span the operational design domain [30], [36], [44]. In this context, an efficientnet-b3 classifier trained and validated on locally representative splits provides reliable upstream signals with strong throughput. Stability vs. severity, misclassification costs for routing errors, and probability calibration are emphasized so the system can avoid or down-weight high-risk adaptations when confidence is low. Integrating uncertainty quantification further improves decisions by separating irreducible data noise from parameter uncertainty and calibrating posteriors [11], [32].

Open issues remain. Condition labels can be coarse or inconsistent across datasets; seasonal and sensor drift can break calibration; and single-label taxonomies struggle with mixed regimes. These motivate hybrid targets (class + severity) and periodic recalibration, along with self-supervised batch-statistics adaptation to slow drift. Regarding model size, with modest labeled sets (20k–30k images), EfficientNet-B3 can generalize and calibrate better than B4 under the same budget, reducing overfitting and variance across folds [23], [25]. B3 also lowers latency and activation footprint without sacrificing class-wise F1 on rain/fog/night in local validations; once data scale/diversity rises, B4's extra capacity may pay off [23], [25].

2.2.2 Object Detection

Object detection reduces complex scenes to actionable entities: vehicles, pedestrians/cyclists, obstacles, and surface defects pertinent to planning and control. Two-stage detectors (e.g., Faster R-CNN) excel at localization in controlled settings, but single-stage designs dominate embedded contexts by removing proposal generation while retaining strong accuracy. YOLOv8 suits robustness–throughput trade-offs with modern label assignment, decoupled classification/regression heads, and efficient backbones enabling real-time inference on constrained hardware.

Anchor-free formulation predicts positions and box parameters directly on dense feature maps, avoiding large prior grids and easing training under distribution shift (adverse weather perturbs object scales/aspects). It also improves small-object sensitivity by eliminating anchor competition. Decoupled heads let classification and regression adopt task-appropriate normalization/receptive fields—useful because weather differentially impacts textures (classification) vs. edges/shapes (localization). Task-Aligned Learning (TAL) dynamically matches predictions to targets by combining classification scores with localization quality, mitigating noisy assignments when low contrast or blur corrupts IoU in fog/rain.

Loss design matters in adverse conditions. Modern YOLO recipes combine classification (BCE/focal) with overlap-centric localization (IoU/GIoU/DIoU/CIoU) and shape/distance terms to stabilize centers and aspect ratios when visibility is poor. Confidence calibration (e.g., temperature scaling) helps produce probabilistically meaningful scores for risk-aware planners. Comparisons against YOLOv5, YOLOX, and two-stage baselines show YOLOv8's anchor-free design and improved matching enhance small/occluded robustness at similar latency; detached heads and training recipes yield stable convergence on mixed-condition corpora. In South Asian scenes, YOLO-family models trained with region-specific augmentation maintain accuracy under illumination/clutter changes and tight budgets, particularly with condition-aware training [30], [44]. Industry reviews concur that dataset realism, condition stratification, and calibration often deliver larger deployment gains than small architectural tweaks [23], [32].

Powerful deployments emphasize schedules of SGD/AdamW with cosine/step LR, mixed precision training, photometric augmentations (exposure changes, color casts, blur, noise) and geometric transforms. Condition stratified validation (clear, rain, fog, low-light) shows clusters of failure covered by global mAP. The compute cost will be calculated as end-to-end latency with improvements/routing and not detector run time. Limitations: bad weather induces fuzzy edges and long-tail small/occluded objects, motion blur/scattering physics may be corrupted by naive enhancement, higher mAP is not safer, etc. Upstream condition-wise co-tuning (and explicit calibration curves) are necessary to expose confidence under shift.

2.2.3 Object Detection in Adverse Weather

Poor weather is a two-fold challenge, i.e., input degradation (loss of discriminative cues) and train-test distribution drift. Restoration has been demonstrated to aid recognition when edges, textures and colors are preserved; over-smoothing or color drift may be detrimental to recognition as opposed to raw inputs [9]. Learned perceptual-goal multi-scale-prior deraining/desnowing reviews are also desirable over pure pixel fidelity to downstream detection [12]. In low-light literature, transforms that are exposure-conscious are focused on to address noise/contrast collapse [15]. Trained single enhancement modules across regimes can increase consistency in cases where validation is achieved by examining beyond reconstruction scores to downstream perception [43].

Detector-internal strategies complement (or substitute) preprocessing when latency/sensor noise constrain pipelines. Reports on detection under adverse conditions highlight normalization choices, receptive-field tuning, and augmentation aligned with image formation, showing measurable drops in false positives and improved localization stability [24]. Weather-condition surveys recommend physics-grounded enhancement over aesthetic filters, and selective use of deformable convolutions [28]. Curricula that start on clear scenes and gradually increase degradation stabilize convergence and prevent overfitting to a single regime. A recurring result is that enhancement should be treated as part of the perception system: enhancement and detection are best tuned jointly so enhancer output statistics align with detector feature sensitivities [17], [43]. Local studies show YOLO detectors fine-tuned with regional data and evaluated on condition splits are more stable in summer rain and winter haze, especially with condition-specific normalization and realistic augmentation [30], [44]. Caveats include enhancer artifacts under noisy/dirty optics, brittle discrete routing, and lab–field gaps. Training detectors on both raw and enhanced streams (dual-path validation) and using robustness-driven augmentation help. Where possible, backup sensing or geometric priors maintain stability when enhancement is unreliable.

2.2.4 Attention Mechanism in YOLOv8 for Detection

Recent works on object detection in bad weather have focused more and more on incorporating attention mechanisms into YOLOv8 architectures for improved feature extraction and robustness [21], [41]. Of these, the Convolutional Block Attention

Module (CBAM) has been particularly effective, improving the detection performance in difficult weather conditions by allowing models to concentrate on salient features while suppressing less relevant information through channel and spatial attention mechanisms [4], [21].

The EMA module into YOLOv8's backbone can be used to fuse contextual information across various scales and enhance pixel-level attention to high-level feature maps with a focus on the detection of bad weather conditions [21]. More recently, the authors presented YOLOv8-STE, a model which employs a Swin Transformer (ST) feature extraction layer in the backbone to enable global interaction and aggregation of the feature maps, along with an EMA mechanism in the neck network that captures feature information at different spatial scales by utilizing multi-scale attention [41]. Another example of the effectiveness of embedding CBAM is demonstrated for the framework YOLOv11-TWCS, with the inclusion of the attention module after each C3k2 layer to highlight salient features while suppressing irrelevant information, reaching 59.1% mAP@0.5 with minimal computation overhead[20].

Comparative studies have shown that attention mechanisms significantly improve detection accuracy in bad weather condition. Moreover, RFCS-YOLO has demonstrated how the integration of CBAM into cross-scale fusion modules comprehensively solves challenges such as complex backgrounds and problems of missing and misdetecting traffic targets in bad weather conditions [37]. Parallel insights are drawn from the underwater detection research, where integration of CBAM on the last layer of the YOLOv8 backbone network enhanced model perception and generalization capabilities of the model in low visibility. Which is also similar to that caused by fog and rain conditions in autonomous driving [29]. Moreover, hybrid schemes incorporating CBAM with transformer encoder blocks in a CSPDarknet53 backbone have revealed better performance in capturing comprehensive global information and contextual information, particularly effective in densely packed environments with degraded visibility [27].

The strategic placement of attention modules within YOLOv8's architecture has been critical in performance optimization. Research proves that integrating CBAM before the Spatial Pyramid Pooling Fast (SPPF) module within the backbone network can fully leverage attention advantages by enhancing feature focus before multi-scale pooling operations. This in turn performs very well in low-contrast and high-background-noise environments [42]. Similarly, the C2f module in YOLOv8, replacing the C3 module from YOLOv5. That which has lower computational requirements while substantially improving convergence speed and overall performance, thus the best place to integrate an attention mechanism [39]. This architectural evolution motivates the approach taken in the present work-integrating CBAM after the last C2f block in YOLOv8m's backbone to improve detection robustness. This is cause Bangladesh-based driving scenarios are frequently exposed to adverse weather conditions that significantly worsen the quality of visual perception.

2.2.5 Bottleneck Classifiers

Model efficiency has become a central concern in modern deep learning, particularly when adapting large pretrained networks to resource-constrained environments or specialized domains. Instead of retraining entire backbones, recent studies advocate for **parameter-efficient adaptation**—freezing most convolutional layers and optimizing lightweight bottleneck classifiers that retain accuracy while reducing trainable parameter counts. Such approaches preserve the representational power of pretrained feature extractors while minimizing computational and memory overhead.

Qian et al. presented *AllWeather-Net*, which employs a compact enhancement head for unified low-light and adverse-weather adaptation without retraining full encoder layers. Keh et al.'s dual-learning design also employed bottleneck projections to efficiently fuse weather-conditioned representations during detection [26]. Tan et al. demonstrated that end-to-end weather adaptation could be achieved with limited fine-tuning when using parameter-efficient intermediate layers and decoupled classifier heads [35].

In object detection, Fang et al. achieved computational savings by restructuring YOLO's detection head with reduced-channel convolution layers optimized for foggy environments [14]. Similarly, Wang et al. showed that bottleneck-based joint restoration–detection networks could maintain high accuracy in snowy conditions with fewer tunable weights [38]. These methods reinforce the architectural logic behind this work's modified EfficientNet-B3 bottleneck classifier: a frozen backbone coupled with a compact fully connected head ($1536 \rightarrow 256 \rightarrow 4$) and dropout regularization, designed to retain classification accuracy while substantially lowering the number of trainable parameters.

Challenges include poor depth in low-texture regions, failures under intense glare/wetness that violate monocular assumptions, and added latency from multi-sensor/multi-task stacks. Mitigations include confidence-gated activation, depth-prioritized regions to bound compute, and semantic gating to disable expensive modules in low-complexity scenes.

2.2.6 Synthetic Data Generation

Adverse-case scarcity in target domains motivates both physics-based and generative synthesis. Unpaired image-to-image translation (CycleGAN) is effective when paired captures are infeasible; two generators and discriminators, trained with adversarial and cycle-consistency losses (plus identity loss), constrain mappings to preserve color/brightness when domains already match. Practically, CycleGAN both normalizes adverse inputs toward detector pretraining statistics and augments clear local captures with realistic adverse variants to reduce weather-induced shifts. Paired Pix2Pix normalization has improved downstream detection when trained with perception-aware objectives [36]. Residual-block generators (6/9 blocks) with strided encoders and fractionally-strided decoders maintain mid-level structure critical for weather transformations; PatchGAN discriminators emphasize high-frequency realism (streaks, haze texture, noise). For driving weather, residual generators are preferred over pure U-Net variants to avoid leaking source-domain artifacts into targets [17], [18], [43].

Physics-based rendering complements generative methods with controllable degradations tied to image formation: atmospheric scattering models for haze, stochastic streak kernels and motion blur for rain, and exposure/noise models for low light [9], [12], [15]. While synthetic outputs cannot fully capture fine-grained South Asian appearance, mixtures of physics-based renders, CycleGAN variants, and in-the-wild captures better cover rare/compound scenarios. Reviews recommend optics-grounded transformations for transferability to field conditions [18], [43]. Mitigations against texture hallucination/color drift include identity/perceptual constraints, mild-to-severe curricula, domain randomization of generators, and detector-in-the-loop validation prioritizing perception metrics over pixel scores [17], [43].

2.2.7 Transfer Learning

Weather-invariant perception relies heavily on transfer learning because collecting large, balanced labels for every target environment is impractical. Typical practice pretrains on large source corpora and fine-tunes on target data. Freezing early layers stabilizes early adaptation; deeper layers are unfrozen as target gradients stabilize. Regularization (weight decay, label smoothing) reduces overfitting to small local sets. Optimizers include SGD with momentum or AdamW, with cosine/step-decay scheduling and mixed precision for throughput. Photometric augmentations (exposure drift, color cast, sensor noise) plus geometry transforms better anticipate adverse conditions than purely aesthetic edits.

A small amount of high-quality local data can close much of the gap between foreign pretraining and domestic deployment, provided validation is condition- and class-stratified so rare, safety-critical gains are not averaged away [18], [23]. Layer-wise unfreezing schedules and condition-aware early stopping prevent catastrophic forgetting while fitting target styles. Selective reweighting/class balancing addresses structural biases in Western corpora (lane-marking assumptions, vehicle-shape priors) [18], [23]. Photometric enhancements aligned with image-formation effects transfer more robustly than cosmetic edits, and curriculum-style difficulty schedules stabilize convergence in adverse regimes [18]. Risks include bias amplification with small/imbalanced local sets, augmentation choices that misrepresent image formation, and overfitting from overly aggressive unfreezing. Tracking expected calibration error and applying temperature scaling preserves probability accuracy over seasonal/sensor drift [11], [32]. Memory limits are mitigated via mixed precision and gradient accumulation; if domain shift exceeds backbone capacity, scaling (e.g., EfficientNet-B3$\rightarrow$B4) can help at larger data scales, while smaller backbones may generalize better $\leq$80k images [23], [25].

2.2.8 Domain Adaptation

Domain adaptation reduces the residual train–test gap without full labeling. Self-training with confidence-filtered pseudo-labels increases coverage; unpaired translation aligns styles; test-time adaptation updates normalization statistics or minimizes entropy online as illumination and precipitation shift. Recent work emphasizes representation-level tuning to handle feature-space shifts, simplifying maintenance relative to continuous pixel-space edits [35]. Lightweight online risk estimators based on batch-norm statistics compensate for slow scene drift in streaming.

Pitfalls include confirmation bias in pseudo-labels, instability in adversarial alignment, and catastrophic forgetting of base domains. Remedies include conservative confidence thresholds, temporal ensembling, mixed-domain rehearsal during adaptation, and periodic evaluation on held-out clear sets. In practice, adaptation frequency is rate-limited to avoid oscillation in rapidly changing scenes and is gated by condition classifiers to signal when adaptation is appropriate. These safeguards let monitors/classifiers coordinate through clear↔cloudy, day↔night, and seasonal transitions.

2.2.9 Uncertainty Quantification

Risk-aware autonomy in adverse conditions depends on reliable uncertainty estimates. Aleatoric uncertainty arises from irreducible data noise (sensor noise, motion blur); epistemic uncertainty stems from limited or shifting training data. Methods such as deep ensembles, Monte Carlo dropout, and evidential predictors—well studied in geosciences/ML—translate to autonomous sensing for estimation and calibration [11]. In deployment, calibrated uncertainty gates enhancement strength, modulates platform speed, and defers decisions when confidence is low. For planners that consume detector confidence, predicted probabilities must match empirical frequencies; calibration via temperature scaling, isotonic regression, or focal variants is common [11], [32]. Practical guidance includes coarse but stable estimates, periodic recalibration with fresh data, and condition-stratified reliability diagrams so planners get state-specific trust levels.

2.2.10 South Asian Deployment Context

South Asian deployments differ markedly from Western benchmarks in weather, infrastructure, and traffic. Monsoons bring prolonged heavy rain with specular films and streaks; winter haze/fog depress visibility and contrast, compounding low-light conditions. Traffic mixes are diverse (rickshaws, motorcycles, vendors, animals, pedestrians) in narrow, mixed-use corridors. Surfaces on roads are usually irregular, there are potholes, temporary obstructions, low/no lane markings, and irregular signs. These influences modify the patterns of occlusion, object scale, appearance on cues as compared to ordered multilane arteries common to Western corpora.

Localized theses fill this lacuna. Work on Bangladesh demonstrates that U-Net road segmenters, which are trained on local imagery, can correctly identify the driveable areas; YOLO family detectors, which are trained using realistic local augmentation, can better detect objects in low light and rain [44]Hazard-detecting smartphone cameras (e.g., bumps/potholes) trained on a diversified illumination achieves better performance compared to models trained on foreign data, which highlights the importance of data realism over architecture [30]Local weather/illumination splits are better perceived and detected with paired Pix2Pix normalization [36] The depth-

sensitive region prioritization computationally spares with limited consequences on the quality of near field detection- useful in embedded applications [33]. Assistive navigation pipelines combining detection, semantics, and language grounding demonstrate transfer of weather-robust perception beyond AVs [34].
These findings emphasize matching data characteristics to the operational domain, condition-aware training/validation, and augmenting appearance models with geometry/semantics. Regulatory, infrastructure, and cultural factors (e.g., variable lighting in mixed-use corridors, inconsistent right-of-way) further recommend conservative planning and uncertainty-sensitive calibration to preserve safety margins under failure.

2.2.11 Research Gaps

Open gaps remain. (1) Lack of South Asia–centric compound-condition benchmarks hinders comparability and external validity [5]. (2) Enhancement and detection are often optimized separately; fully joint training for perception remains uncommon despite advantages of perception-guided enhancement [17], [43]. (3) Perception metrics seldom link to driving outcomes (e.g., intervention rates, TTC margins) across weather bands, forcing reliance on proxy metrics [23], [32]. (4) Compute-aware resiliency methods that preserve near-field recall under latency constraints are under-explored [24]. (5) Condition-wise calibration and routine uncertainty reporting are not standard despite their importance for risk-aware planning [11], [32]. (6) Many enhancement methods are not physics-grounded, limiting field transferability [18]. (7) Domain adaptation is vulnerable to catastrophic forgetting and confirmation bias; conservative, stable streaming protocols are still evolving [35]. (8) Cross-modal fusion lacks standardized alignment/reliability indicators for adverse conditions, limiting reproducibility and deployment guidance [22], [31].

2.3 Emerging Design Patterns and Thesis positioning.

The literature converges on several patterns. Perception-sensitive enhancement is favored over generic reconstruction: restoration modules are selected/trained to preserve discriminative structure and are evaluated on downstream detection, not just reconstruction [9], [15], [17], [43]. Real-time one-stage detectors are trained with augmentation aligned to photometric physics and are calibrated to maintain correct-confidence under shift [18], [24], [28], [32]. Under low visibility, lightweight geometry/semantics and region prioritization focus compute on highest-risk areas [33], [44]. An upstream efficientnet-b3 reliably informs experts or sets parameters for a single enhancer [16], [25], [43]. Data engines mix unpaired translation, physics-based rendering, and in-the-wild captures to cover adverse regimes, while adaptation and uncertainty estimates sustain performance and expose planner confidence.

We fine-tune EfficientNet-B3 on a new locally curated dataset for environment classification and evaluate YOLOv8 under Bangladesh rain/fog/night. Unpaired CycleGAN provides adverse$\leftrightarrow$clear normalization and clear$\rightarrow$adverse augmentation; physics-based rendering supplies controllable severity for curricula/ablations.

Lightweight geometric cues (depth-aware region prioritization) are used where helpful to improve near-field recall without the cost of full 3D sensing. Transfer learning is combined with domain/test-time adaptation to stabilize performance across seasons and diurnal cycles. Uncertainty-aware validation ensures confidence aligns with operational risk. This synthesis reflects robust patterns in restoration, detection, data generation, adaptation, and calibration [1], [43], supported by localized theses demonstrating that data realism and pipeline integration dominate architectural tweaks for reliable deployments on South Asian roads [30], [33], [34], [36], [44].

Chapter 3

Requirements, Impacts and Constraints

3.1 Final Specifications and Requirements

Baseline dataset images were taken with a camera device attached to a dashcam mount fixed on the windshield of a vehicle. It is equipped with a 48MP primary camera with f/1.78 aperture, sensor-shift optical image stabilization, and a 24mm-equivalent focal length. Images were taken with a resolution of 4032×3024 with automatic exposure, HDR on, and computational photography with active features to achieve natural color rendition and dynamic range typical of common consumer-grade automotive cameras. The mount angle was fixed with a forward view about 1.5 meters above ground, with horizontal alignment and small pitch angle to achieve typical driver perspective and consistency in data collection routes.
All experiments were conducted on a workstation equipped with an NVIDIA RTX 3080 GPU, 64 GB of system RAM, and a 12-core CPU. The system operated under customized Conda environment with Python 3.10, PyTorch 2.1, and CUDA 12.1. For reproducibility, all random seeds were fixed, and library versions were pinned in a requirements file.

Training environment. Training was executed on a single GPU using mixed precision to accelerate computation. The dataset was preprocessed using OpenCV and torchvision utilities, and all experiment configurations were logged with timestamps and model checkpoints.

Hardware efficiency profiling. Inference latency and memory usage were measured for both classification models to demonstrate deployment feasibility on embedded hardware. Timing was averaged over 500 frames with batch size 1 to simulate real-time streaming.

3.2 Societal Impact

This research addresses prominent safety and access concerns in autonomous and semi-autonomous driving systems, particularly for underserved regions in world datasets. By developing weather-informing perception models trained in locally appropriate driving scenarios in Bangladesh, it facilitates more equitable rollouts of autonomous vehicles in South Asia and new regions. Stable perception in adverse weather can reduce accidents, improve traffic flow, and make transport infrastructure safely deployable in countries where weather conditions and roadway topography deviate from Western norms. Moreover, the dual-task assessment protocol offers a clear methodology for evaluating models' strength, thereby facilitating-informed choice among engineers and policymakers for autonomous systems' implementation. Open nature of methodologies combined with lessons from such works necessitates greater computer vision research investment by scientific communities in developing countries, thereby enhancing local innovativeness while further diminishing foreign dataset and model dependency. But societal risk involves risk of excessive reliance on automated systems in scenarios where they remain untested, and risk of biased operation in case the dataset fails to encompass the full range of local driving habits and variations in local infrastructure. Validation with caution and phased deployment is necessary to counter these risks.

3.3 Environmental Impacts

The environmental cost of this project is comprised of computational power use in training and inferring models, as well as indirect ramifications of enabling autonomous vehicle technologies. Training deep neural models, particularly in hyperparameter search and multistage fine-tuning, requires significant use of the GPU. Our experiment consumed around 120 hours of cumulative GPU time for all combined experiments. Using an NVIDIA RTX 3080 with an average power of 320W, it corresponds to around 38.4 kWh of direct energy usage, or around 19 kg of CO emissions with an assumed grid carbon intensity of 0.5 kg CO/kWh (typical of mixed energy grids). From a positive standpoint, advanced perception systems can potentially achieve ecological benefits by enabling smoother traffic flow, reducing unnecessary braking and acceleration, and enabling electric and autonomous vehicles that can increase energy efficiency. Furthermore, weather-aware models may reduce the need for unnecessary sensing hardware (including expensive LiDAR systems), thus cutting material and energy costs associated with vehicle production. Future research should also emphasize the use of effective training techniques, e.g. transfer learning, premature termination, and model distillation in minimizing the environmental effects. Also, implementing simplified models in edge devices instead of cloud-based inference can save energy expenses in data transmission and render it viable in sustainable autonomous system structures.

3.4 Ethical Issues

There are numerous ethical concerns in this research. The first one is that of data privacy; ground truth dataset has been gathered in the open areas and is anonymous, but it has been taken into account not to include frames containing recognizably familiar faces or license plates. It was also conducted to make sure that the privacy laws were adhered to, and the data collection was not performed in prohibited or personal places and was done through a detailed examination of all the images. Second, there is the question of prejudice and impartiality: the dataset reflects Dhaka-based driving situations, which may not be applicable to countryside or different cities or location in Bangladesh.This geographical particularity may be induced to cause biased model behavior when an extrapolation is done. In future research, areas of focus should include enhancing the data collection to guarantee inclusion of heterogeneous population, road types, and type of traffic. Thirdly, the safety and accountability matters are important: even though the given study provides the comprehensive framework of evaluation, the models designed in this work are not certified to be safe in terms of the use of the real-world autonomous vehicles. Poor application of these models in the absence of rigorous validation can lead to safety related accidents. Comprehensive records that highlight restrictions and planned uses are provided to prevent use in a wrong manner. System integrators and manufacturers must bear the responsibility of making deployment decisions, and they must conduct validation that is relevant in areas particular to them, and comply with the regulatory requirements of the area they are working in. In conclusion, dual-use issues must be acknowledged: although the main objective of this research is to enhance the safety of transportation, perception technologies may additionally be utilized in surveillance or military scenarios. The authors emphasize the importance of responsible application that aligns with societal welfare and caution against uses that may violate civil liberties or human rights.

3.5 Risk Management

The main technical risks present in this research are model overfitting, dataset bias, and failure of generalization in unforeseen circumstances. To prevent overfitting, aggressive dropout regularization (0.5 and 0.8), as well as progressive unfreezing of backbone networks, were carried out during training. Cross-validation and early stopping by means of validation loss minimized the risk of memorization additionally. Dataset bias was managed by stratified splitting based on route and time-of-day for balanced representation of all weather classes and scene contexts. The geographical restriction to Dhaka, however, brings about the risk of failure of generalization; it is advised to validate on external datasets prior to deployment. Operational risk also entails deployment in new weather situations (e.g., intense storms, sandstorms, or mixed type rain/snow precipitation) with unclear model behavior. To handle this, we suggest confidence-based thresholding: systems must alert low-confidence predictions and default to human operators or redundant sensors when doubts about weather type exceed safety limits. Also, periodic model recalibration with new data is necessary to address sensor drift, seasonal variations, and changing traffic conditions. The project management risk was reduced by iterative development, code design in modular form, and version control (Git). Computational resource problems were addressed by making use of mixed-precision training as well as data pipelines that were efficient. Third-party library dependency problems were addressed through version pinning as well as creating reproducible environments.

3.6 Economic Analysis

The economics of such research is framed by development cost, deployment cost, and return on investment by increased safety and greater efficiency. Development expenses: Major costs involve computational hardware (GPU workstation), data collection (fuel, time, access to a vehicle), and human labor (researcher time for annotation, training of models, verification). If we also assume that the research would take up six months with joint access to available university equipment, then the direct cost is estimated to run to around $2,000$–3,000 (including power, cloud compute credits for backup experiments, and incidental expenses of hardware). Using pretrained models and open-source software also minimized development expenses by training from scratch. Cost of deployment: Deploying the proposed weather classifier on an edge device (e.g., NVIDIA Jetson Nano or similar) is in the range of 100–300 per unit of vehicles. Inference latency as well as the memory requirements (profiled in Section 3.1) indicate that baseline as well as custom models would run on mid-range embedded hardware, not requiring expensive accelerators. This renders the solution affordable for mass-market integration in assisted drive systems or cheaper autonomous vehicles for new markets. Return on investment: Enhanced reliability of perception in bad weather can lower accident rates, decrease insurance premiums, and reduce car downtime as a consequence of weather-related accidents. For ride-hailing or fleet operators in areas with typical monsoons or fogs, incremental advances in detection accuracy can amount to substantial financial savings in terms of lowered liability as well as enhanced continuity of operation. The research's methodological advancements, especially the dual-task evaluation framework, can also help bring down the time-to-market for weather-robust perception systems by implementing validated baselines as well as dataset augmentation pipelines that can be reused in the future, thus accelerating upcoming RD. Scalability: Incremental deployment is facilitated by the framework's modular design: classification of weather can be incorporated as a low-cost diagnostic system in isolation, while detection robustness assessment can inform procurement of sensor suites. Scalability facilitates phased adoption in accordance with budget limitations and timelines of regulation. On balance, the economics of this project are favorable for academic, commercial, and public use, particularly in low-priced markets where redundancy of sensors and infrastructure is minimal.

Chapter 4

Methodology

4.1 Overview of Proposed Approach

This chapter presents a weather-aware object detection framework designed to quantify how adverse weather affects perception in urban driving scenarios relevant to Bangladesh. The framework consists of two components trained on identical inputs: (i) a weather classification model based on EfficientNet-B3 and its modified bottleneck derivative (our custom model), and (ii) a CBAM-enhanced YOLOv8m object detector initialized with pretrained weights and fine-tuned for adverse weather robustness.The design decision that has been made is that the attention mechanisms should be directly designed to be part of the detector backbone so that they can be used in recalibrating features in case of unfavorable weather. To be more precise, another Convolutional Block Attention Module (CBAM) is added at the end of the C2f block of the backbone of the YOLOv8m, which enables the model to focus attention on salient features selectively, rather than focusing on the noise produced by weather conditions. This architectural improvement is accompanied by a classification of weather conditions that allows stratified performance analysis in varying weather conditions and gives interpretable signals to comprehending the detection behavior in various weather conditions.

On a high-level, the pipeline will go through the following. The scenes of clear weather are gathered to create a base corpus. The individual scenes are then complementarily augmented into three bad-weather conditions (fog, rain, snow) with the help of complementary augmentation modules. The resulting dataset is split into training, validation, and test partitions under route and time-of-day stratification. EfficientNet-B3 is fine-tuned for weather classification (four classes: clear, fog, rain, snow) using a two-stage curriculum (frozen backbone, then progressive unfreezing). Our modified model is trained with a matched curriculum, furthermore, adding strategic feature compression, selective layer unfreezing, and calibrated regularization, which provides lower inference time while preserving accuracy. Both baseline YOLOv8m and CBAM-enhanced YOLOv8m are trained on the dataset with properly annotated bounding boxes. We compute classification metrics for the weather model(s) and compare detection metrics (mAP, precision, recall) between baseline and CBAM-enhanced variants, stratified by weather class to quantify the attention mechanism's effectiveness under different environmental conditions. Finally, we synthesize the results to quantify how CBAM integration improves detection precision/recall under adverse weather and to articulate the computational efficiency–accuracy trade-offs for deployment on resource-constrained autonomous systems.

Design constraints and requirements. The design targets near real-time inference on modest GPUs or edge accelerators; memory is constrained, and reproducibility is prioritized (fixed seeds, deterministic ops where possible). Both detector variants (baseline and CBAM-enhanced) are trained on the annotated Bangladesh-specific dataset to enable fair comparison of architectural modifications under local weather conditions. The classifier must be robust across day-time conditions, mixed traffic, and variable scene composition, and its confidence must correlate usefully with detection stability to support uncertainty-aware policies.

4.2 Dataset Development

We describe the base collection protocol, quality assurance steps, splitting strategy, augmentation modules, and final dataset composition.

4.2.1 Base Dataset Collection

Collection protocol. The base corpus comprises 3,990 *clear-weather* images captured from Dhaka city traffic. The collection plan emphasized diversity: multiple arterial and local routes, intersections and mid-block scenes, mixed traffic densities (off-peak, peak). The sampling cadences and capture intervals used were selected such that successive frames were not trivially redundant, thus making the maximum use of scene variation on the resource budget.
Stabilized optic cameras were mounted on cars at the windshield level to simulate the perspective of the front-facing car detectors. The common set of parameters was full-HD or higher resolution, 30 fps (images subsampled off the video to prevent excessive temporal correlation), a medium field of view that offers a good balance between scene coverage and geometric distortion. Hardware was mounted to reduce vibration; it was held approximately horizontal with a small tilt downwards to record road and near-field interactions. Clock / timezone consistent Time-synchronized logging: Time-of-day stratification was reliable.
The routes were selected in order to cover the central business districts and residential areas and the arterials between the major hubs. It was aimed at the representative combination of signs, lane marking, curbs space usage, and pedestrian flows. Notably, the collection was done in extremely congested regions (where they are valuable) and the regions with greater speed (where distant objects are prevalent). The mixture is essential in assessing visibility-restricted circumstances like fog and snow.

Annotation strategy. Image-level weather labels are used for training and evaluation of the classifier. Additionally, bounding boxes for eight object classes (Person, Traffic signal, Bus, Car, Rickshaw, Taxi, Motorcycle, Pothole) were manually annotated using Roboflow to enable detector training and evaluation across weather conditions. Clear images were manually confirmed, and quality screens (below) removed frames with excessive blur, exposure extremes, or privacy concerns (faces/plates unmasked) prior to augmentation.

4.2.2 Data Organization and Splitting

We split the 3,990 clear images into train/validation/test subsets at 70/15/15 proportions. Splitting is stratified by route and time-of-day: each stratum contributes samples proportionally, preventing leakage where nearly identical scenes appear across partitions. After augmentation, the same split indices are applied to each weather class to preserve cross-weather alignment of scenes. This enables paired analysis: a single base scene yields four variants (clear/fog/rain/snow) that live in the same partition and can be compared without confounds.

4.2.3 Weather Augmentation Pipeline

We generate three adverse-weather variants per clear image using complementary modules designed to model different aspects of the image formation process: physically motivated atmospheric attenuation for fog, and learning-based texture synthesis for rain and snow.

Physics-Based Fog Synthesis

Depth map generation methodology. Monocular depth estimation provides a relative depth map $d(x)$ for each pixel x. We use a state-of-the-art encoder–decoder (e.g., DPT/MiDaS-style) that generalizes well to urban driving scenes. Depths are normalized to $[0, 1]$ per image to accommodate relative rather than metric scale, which is sufficient for fog transmission synthesis.

Atmospheric scattering model. We adopt the standard Koschmieder model for haze/fog:

$$I(x) \;=\; J(x)\,t(x) \;+\; A\,\big(1 - t(x)\big), \qquad t(x) \;=\; \exp\big(-\,\beta\,d(x)\big), \tag{4.1}$$

where $J(x)$ is the clear scene radiance, $I(x)$ is the observed foggy image, A is global airlight (per-channel constant), $d(x)$ is the estimated depth, and $\beta > 0$ is the scattering coefficient controlling density. This model captures two effects: distant radiance is attenuated and replaced by airlight, producing the characteristic veiling luminance that reduces contrast.

Fog density parameters and variation. We vary (A, β) across a grid to synthesize light, medium, and heavy fog levels. Airlight A is sampled in $[0.7, 1.0]$ (normalized intensity), while β spans small to moderate attenuation values, e.g., $\beta \in \{0.05, 0.15, 0.25\}$. Sampling is done per image with stratified draws to ensure class balance across intensities.

Perceptual safeguards. To avoid artifacts, we clamp transmission $t(x)$ to a minimum floor so highlights are not over-suppressed, and we estimate A using bright sky/upper-image regions to anchor plausible airlight colors. We also avoid excessive channel imbalance in A so fog appears neutral unless the ambient lighting strongly suggests color shift.

GAN-Based Rain and Snow Augmentation

Architecture details. Rain is characterized by directional, semi-transparent streaks and dynamic scene effects (wet surfaces, local blur). We use a ResNet-based generator with residual blocks and a CycleGAN discriminator. The generator learns to overlay streak textures and induce mild motion-blur-like effects where appropriate, while preserving scene geometry.

Training procedure. If paired clear/rain data is unavailable, an unpaired translation setup (Cycle-like constraints) is used with adversarial loss, cycle-consistency (if applicable), and perceptual loss on high-level features to preserve semantic content. The discriminator encourages realism of local patches. We control intensity via generator inputs or feature scaling, yielding snow and rain variants distinguished by streak density, length, and blur.

Intensity control and realism checks. Parameters govern streak orientation coherence, per-patch density, and spatial heterogeneity (e.g., stronger near the camera, weaker in distance). We avoid spatially uniform rain which looks artificial; local variations and wet surface highlights are introduced so that rain signals align with physics (vertical bias with mild wind-induced slant, specularities on road).

Manual Object Annotation

Annotation platform and workflow. Manual bounding box annotations were performed using Roboflow, a cloud-based annotation platform that provides efficient labeling tools, quality assurance features, and export capabilities compatible with YOLO format. The annotation process followed a systematic workflow: image upload, class definition, bounding box drawing, quality review, and export in YOLOv8-compatible format.

Class selection and rationale. Eight object classes were defined to capture the most safety-critical and prevalent entities in Bangladesh urban driving scenes:

- **Person:** Pedestrians, including those crossing streets, waiting at intersections, or near roadways—critical for collision avoidance.
- **Traffic signal:** Traffic lights and signals that govern intersection behavior and right-of-way decisions.
- **Bus:** Large public transport vehicles common in Dhaka's arterial routes, requiring early detection due to size and stopping patterns.
- **Car:** Private passenger vehicles representing the majority of four-wheeled traffic.
- **Rickshaw:** Three-wheeled cycle rickshaws unique to South Asian traffic, often unpredictable in movement.
- **Taxi:** Cars which are used commercially as ride-sharing cars or taxis.

- **Motorcycle:** Two-wheel powered vehicles that are likely to lane-squeeze and turn fast.
- **Pothole:** Road surface anomalies that are dangerous to the stability of vehicles and involve changes in the path planning.

The set of classes balances detection of vulnerable road users (Person), traffic control in infrastructure (Traffic signal), and various types of vehicles - they represent the local traffic composition (Bus, Car, Rickshaw, Taxi, Motorcycle); and road hazards (Pothole).

Annotation guidelines and quality control. The guide lines were strictly adhered to in order to have uniformity.

- **Bounding box tightness:** Boxes were drawn to tightly enclose the visible extent of each object, minimizing background inclusion while capturing the full object silhouette.
- **Occlusion handling:** Partially occluded objects were annotated if at least 30% of the object was visible and the class was identifiable. Cases that were overly occluded were also removed to prevent uncertain training cues.
- **Truncation policy:** Objects extending beyond image boundaries were annotated up to the visible edge, maintaining bounding box validity.
- **Small object threshold:** Objects smaller than 10×10 pixels were excluded due to insufficient resolution for reliable detection.
- **Crowded scenes:** In dense traffic scenarios, all visible instances of each class were annotated to capture realistic object density distributions.

Quality assurance entailed multi-pass review: first annotation, second annotator review and final review to iron out discrepancies. Monitors of inter-annotator agreement were done on a 10 percent subset with the differences debated and settled by consensus.

Annotation statistics and distribution. All 15,940 images were annotated on the four weather conditions (clear, fog, rain, snow) and object labeling was the same irrespective of the visibility degradation. Annotation distribution indicates the natural occurrence of the classes in the urban traffic of Bangladesh:

- **Person:** Some annotations (around 25%), are indicative of heavy pedestrian traffic.
- **Car:** The prevailing type of vehicle is about 20% in value.
- **Motorcycle:** It is about 18%, and this is common traffic in Dhaka.
- **Rickshaw:** There is about 15% of unique regional context.
- **Bus:** About 8% of this market is focused on arterial routes.

- **Taxi:** Taxi is around 6% with the main location being commercial areas.
- **Traffic signal:** There is approximately 5%, limited by the density pd intersection.
- **Pothole:** Around 3%, which is an issue of infrastructure.

This distribution facilitates class-balanced training by weighted sampling strategies and offers enough examples per-class to use in strong detector learning.

Export format and integration. The output of Roboflow was annotations in YOLOv8 format: normalized bounding box coordinates (center x, center y, width, height) with class indices. The export preserved the train/validation/test split stratification of Section 4.2.2, , whereby the annotated objects of each weather variant were consistent with the corresponding partitions. The format was used to directly ingest into YOLOv8 training pipelines without any further preprocessing.

4.2.4 Final Dataset Composition

Applying the above pipeline yields ∼15,940 images across four weather classes (clear, fog, rain, snow), with near-balance by design. All images were manually annotated with bounding boxes for eight object classes using Roboflow, resulting in a fully labeled dataset for both weather classification and object detection tasks. Minor deviations arise from quality filters (e.g., removing a few heavy-fog images that over-saturated. The distribution of samples per class is approximately the same in each split (train/val/test); it facilitates training classifiers and offers fair and stratified detection analysis. The principle of the balance in classes is repeated in the assessment to ensure that the macro and weighted measures are balanced (or no dominant class is present).

4.3 Weather Classification Models

This section presents the weather classification framework, encompassing both the baseline EfficientNet-B3 and a two-layer customized classification head with a 256-dimensional compression bottleneck for this research. Both the models have aim to accurately classify four weather categories—clear, fog, rain, and snow—on a balanced, hybrid-augmented dataset reflecting Bangladesh's driving contexts. The goal is twofold: (1) to establish a strong baseline using EfficientNet-B3, and (2) to evaluate how a **two-layer compression head (1536→256→4)** combined with selective backbone unfreezing and dropout calibration can improve weather-domain adaptation and efficiency..

4.3.1 EfficientNet-B3 Base Model

Architecture overview and motivation. EfficientNet-B3 was selected as the baseline classifier due to its favorable trade-off between accuracy and efficiency achieved through compound scaling of network width, depth, and resolution. The architecture consists of stacked Mobile Inverted Bottleneck Convolution (MBConv) layers with squeeze-and-excitation modules to enable adaptive channel-wise feature

recalibration. The compound scaling coefficients $(\phi, \alpha, \beta, \gamma)$ maintain a balanced growth across dimensions, allowing the model to achieve high representational power while remaining computationally feasible for embedded inference.
The original EfficientNet-B3 head, designed for ImageNet's 1000-way classification, was replaced by a global average pooling layer followed by a fully connected dense layer with four outputs, corresponding to the target weather classes. A softmax activation ensures probabilistic normalization across class logits.

Transfer learning strategy. Weights were initialized from ImageNet-1K pretraining to leverage generic edge, texture, and color pattern representations. The classifier head was reinitialized using Xavier uniform initialization to prevent biased convergence toward pretraining categories. Transfer learning accelerates convergence and enhances generalization, especially when dataset size is limited.

Two-stage fine-tuning protocol. A progressive fine-tuning strategy was adopted to balance stability and adaptability:

- **Stage 1 (frozen backbone):** All convolutional layers were frozen while training only the new classifier head for 10 epochs using a learning rate of 10^{-5}. This allowed rapid adaptation of the head to the new weather task without disturbing pretrained feature hierarchies.

- **Stage 2 (Backbone unfrozen):** The top-level blocks were gradually unfrozen, then mid-level blocks, while reducing the learning rate to 2×10^{-6}, 10^{-6}, 5×10^{-7} to avoid catastrophic forgetting. Validation loss was monitored with early stopping to prevent overfitting.

Training and optimization. Optimization used the Adam optimizer with an initial learning rate of 1×10^{-5}, decayed via a cosine annealing schedule. A batch size of 32 was used given GPU memory constraints. The categorical cross-entropy loss function was minimized over balanced mini-batches. Data augmentation included random brightness, contrast, and hue shifts, along with horizontal flips and mild rotations ($\pm 10°$), providing robustness to lighting and viewpoint variation. Excessive geometric transformations were avoided, as global cues such as fog diffusion and rainfall streaks should remain intact.

Expected performance characteristics. The EfficientNet-B3 baseline is expected to achieve high validation accuracy with relatively slow inference speed and moderate memory usage. Its strong hierarchical representation makes it ideal for benchmarking synergistic efficient variants in subsequent sections.

4.3.2 Custom EfficientNet-B3 (Two-layer classifier design Variant)

Motivation. While EfficientNet-B3 delivers high accuracy, its computational footprint may limit real-time deployment in embedded or on-vehicle hardware. Thus, a custom variant was developed to reduce computational time and floating-point operations (FLOPs) without compromising core representational power. This model,

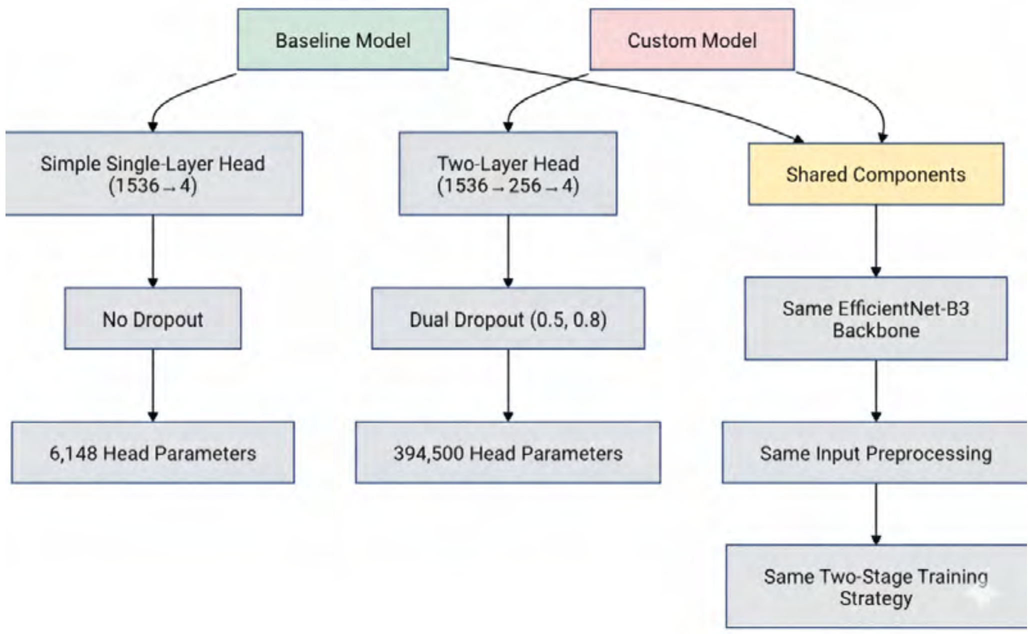

Figure 4.1: Diagram showing the specifications of the Baseline models and the Custom model

referred to as the Custom EfficientNet-B3 variant, employs a two-layer classification head with strategic regularization and progressive training, optimized for improved generalization on weather-specific datasets.

Architectural design. The modified architecture retains the full EfficientNet-B3 backbone and employs a two-stage progressive training strategy. Stage 1 freezes the entire backbone while training only the new two-layer classification head ($1536 \rightarrow 256 \rightarrow 4$) with ReLU non-linearity and aggressive dropout (0.5 and 0.8). Stage 2 progressively unfreezes backbone layers (top 30, then 60, then all layers) with decreasing learning rates (2×10^{-6}, 1×10^{-6}, 5×10^{-7}) to enable controlled domain adaptation while preserving pretrained features. The two-layer head provides enhanced non-linear feature transformation compared to the baseline's single-layer design, while aggressive dropout prevents overfitting on the limited weather dataset. The final classification head is completely redesigned as a two-layer structure: a first fully connected layer ($1536\rightarrow256$) with ReLU activation, followed by dropout layers and a final classifier ($256\rightarrow4$) with softmax activation.
Table 4.1 provides a schematic comparison between the baseline and custom EfficientNetB3 model, based on the proposed two-layer classification head with compression of features, aggressive regularization of dropouts, and tactical scheme of backbone freezing.

Training procedure. The training used the same two stage approach as the baseline though with reduced learning rates at the unfreezing phase ($5 \times 10^{-4} \rightarrow 1 \times 10^{-5}$) so that the gradients are held constant in the smaller model. To stabilize the training with the additional 394K head parameters and avoid overfitting to the small weather data, regularization strength was greatly enhanced with dual dropout layers (0.5 and 0.8).

Table 4.1: Comparison between Baseline and Custom EfficientNetB3 Models

Aspect	**Baseline Model**	**Custom Model**
Head Architecture	Single linear (1536→4)	Two-layer (1536→256→4)
Head Parameters	6,148	394,500
Dropout	None in head	0.5 and 0.8
Backbone Freeze (Stage 1)	100%	100%
Stage 2 Unfreezing	Progressive (30→60→all layers)	Progressive (30→60→all layers)
Non-linearity in Head	No	Yes (ReLU)
Feature Compression	No	Yes (1536→256)
Complexity	Minimal	Moderate

Evaluation metrics. To measure trade-offs, parameter count, computational speed and classification accuracy were compared to the baseline. Although it justified the model design as a balanced complexity architecture to be used with small meteorological datasets, the modified model had higher test accuracy. This enabled objective quantification of whether the model achieved "synergistic efficiency" while retaining weather recognition reliability.

Table 4.2: Hyperparameters for Two-Stage Training

Hyperparameter	**Value**
Stage 1 Epochs	10
Stage 2 Epochs	15
Batch Size	32
Image Size	300×300
Optimizer	Adam
Loss Function	CrossEntropyLoss
Early Stopping	Yes (patience 5/4/4/5)

Summary. This custom variant demonstrates that **selective backbone unfreezing + compression bottleneck + aggressive dropout** forms a synergistic architecture that achieves active domain adaptation — transforming transfer learning from a passive feature reuse strategy into an adaptive weather-domain learner. Its inclusion strengthens the methodological rigor of the dual-task evaluation by validating efficiency-performance trade-offs.

4.4 Object Detection Framework

This section describes the object detection framework designed to evaluate and enhance detection reliability under adverse weather conditions. The framework consists of two architectures: (i) a baseline YOLOv8m model serving as the performance reference, and (ii) a CBAM-enhanced YOLOv8m variant that integrates attention mechanisms. Both models are trained on the annotated Bangladesh driving dataset. Moreover, we also evaluated the models across all four weather classes. Furthermore, we also aim to quantify the effectiveness of attention-based architectural modifications.

4.4.1 Baseline YOLOv8m Configuration

Architecture overview YOLOv8m is an anchor-free single-stage object detection network with real-time inference capabilities. The network's backbone relies upon Cross-Stage Partial modules with two convolution layers (C2f) that allow gradient backpropagation through divided modules. The use of C2f modules has replaced the C3 modules in YOLOv5 and brings improvements in terms of converging faster and requiring fewer calculations while ensuring that the representation capacity remains unaltered. The detection neck incorporates the Path Aggregation Network (PANet) structure with bidirectional feature fusion. The detection head in YOLOv8m decomposes the detection process in both classification and localization objectives. This strikes the perfect balance between detection accuracy and efficiency.

Pretraining and Initialization The baseline model was initialized with COCO-pretrained weights in order to utilize the learned representation of common objects like person and vehicle, which are more transferable in the context of urban driving scenarios. COCO pretraining helps in providing strong initial feature extractors in terms of low-level visual patterns such as edges and textures, and mid-level patterns like object parts and relations in images.

Training configuration. The baseline YOLOv8m was fine-tuned on the annotated Bangladesh dataset using the following hyperparameters:

- **Optimizer:** AdamW with weight decay of 0.0005 to prevent overfitting.
- **Learning rate:** Initial learning rate of 0.001 with cosine annealing schedule, decaying to 0.0001 over training.
- **Batch size:** 16 images per batch, constrained by GPU memory (NVIDIA RTX 3060 12GB).
- **Input resolution:**640×640 pixels, standard for YOLOv8m, balancing detection accuracy and inference speed. and inference speed.
- **Epochs:** 100 epochs and early stopping on validation mAP at 0.5, patience = 10 epochs
- **Data augmentation:** Mosaic (4-image tiling), MixUp (image blending with alpha=0.1), random horizontal flip (p=0.5), HSV color jitter (hue shift ±0.015, saturation ±0.7, value ±0.4), random scaling (0.5–1.5×), and translation (±0.1 of image size).
- **Loss function:** A sum of CIoU (Complete Intersection over Union) loss of bounding box regression, binary cross-entropy of objectness and cross entropy of class prediction.

The split of the training-set (70 percent of all data, about 11,158 images in all weather classes) was split into two parts, and the split was followed by validation to check the validation split (15 percent, about 2,391 images). The checkpoint with the best validation mAP of 0.5 was saved as the baseline.

Inference settings. The parameters of inference which were used throughout the evaluation of experiments were:

- **Confidence threshold:** Should confidence: 0.25, and these predictions are filtered out to minimize false positives.
- **IoU threshold for NMS:** 0.45, which eliminates duplicate detections and maintains individual objects.
- **Maximum detections per image:** Peak number of detections per image: 300 which is sufficient in urban traffic congestions.

These settings were held constant for both baseline and CBAM-enhanced models to ensure fair comparison.

4.4.2 CBAM-Enhanced YOLOv8m Architecture

Motivation for attention mechanisms. The presence of unfavorable atmospheric conditions (fog, rainfall, and snowy images) further incorporates the loss of visibility through various atmospheric phenomena. We see contrast reduction via scattering in the atmosphere. Moreover, we can observe regions hidden by rainfall streaks or regions reduced in visibility due to accumulated snow. Traditional convolutional feature extractors perform calculations uniformly across all spacial locations and feature channels without the capability of adaptive enhancement of relevant areas and corresponding reduction of noisy information caused by unfavorable atmospheric conditions. The attention mechanism involves adaptive recalibration of feature maps.

Convolutional Block Attention Module (CBAM). CBAM is a lightweight attention module that sequentially applies channel attention and spatial attention to refine intermediate feature maps [4]. Given an input feature map $\mathbf{F} \in R^{C\times H\times W}$ (where C is the number of channels, H is height, and W is width), CBAM produces an attention-refined output $\mathbf{F}'' \in R^{C\times H\times W}$ through two sequential stages:
Channel Attention: Focuses on "what" is meaningful by learning inter-channel relationships. The module applies both average pooling and max pooling across spatial dimensions to generate two context descriptors: $\mathbf{F}^c_{\text{avg}} \in R^{C\times 1\times 1}$ and $\mathbf{F}^c_{\text{max}} \in R^{C\times 1\times 1}$. These descriptors are passed through a shared multi-layer perceptron (MLP) with one hidden layer (reduction ratio $r = 16$), and the outputs are summed and passed through a sigmoid activation to produce the channel attention map $\mathbf{M}_c \in R^{C\times 1\times 1}$:

$$\mathbf{M}_c(\mathbf{F}) = \sigma\big(\text{MLP}(\text{AvgPool}(\mathbf{F})) + \text{MLP}(\text{MaxPool}(\mathbf{F}))\big) \tag{4.2}$$

The refined feature map is obtained by element-wise multiplication: $\mathbf{F}' = \mathbf{M}_c \otimes \mathbf{F}$.
Spatial Attention: Focuses on "where" is informative by learning spatial relationships. Average pooling and max pooling are applied across the channel dimension to produce two 2D maps: $\mathbf{F}^s_{\text{avg}} \in R^{1\times H\times W}$ and $\mathbf{F}^s_{\text{max}} \in R^{1\times H\times W}$. These are concatenated and convolved with a 7×7 kernel followed by sigmoid activation to generate the spatial attention map $\mathbf{M}_s \in R^{1\times H\times W}$:

$$\mathbf{M}_s(\mathbf{F}') = \sigma\big(\text{Conv}_{7\times 7}([\text{AvgPool}(\mathbf{F}'); \text{MaxPool}(\mathbf{F}')])\big) \tag{4.3}$$

The final output is: $\mathbf{F}'' = \mathbf{M}_s \otimes \mathbf{F}'$.
CBAM adds minimal computational overhead (less than 1% increase in FLOPs for typical feature map sizes) while providing significant representational enhancement, making it well-suited for resource-constrained autonomous driving systems.

Integration strategy: placement after final C2f block. The CBAM module was strategically inserted after the final C2f block in YOLOv8m's backbone, immediately before the Spatial Pyramid Pooling Fast (SPPF) module. This position was developed on the basis of three factors:

1. **High-level semantic features:** High-level semantic features: Final C2f block is used to generate feature maps that have been subjected to a series of downsampling and fusions to high level semantic features (object shapes, contextual relationships) as opposed to low level edges. At this stage, attention helps the model to focus on object relevant features before multi-scale aggregation.

2. **Moderate spatial resolution:** At this point, feature maps do not lose as much spatial resolution (usually 20x20 with a 640x640 input) to allow meaningful spatial attention. Earlier insertion would operate on higher-resolution maps (increasing computational cost), while later insertion (after SPPF or in the neck) would process heavily downsampled features with limited spatial detail.

3. **Pre-pooling recalibration:** Placing CBAM before SPPF allows attention-refined features to undergo multi-scale pooling (1×1, 5×5, 9×9 kernels in SPPF), ensuring that the most salient features are emphasized during receptive field expansion. This ordering maximizes the utility of attention guidance for subsequent detection head processing.

Implementation-wise, the CBAM module was inserted as a standalone layer in the YOLOv8 architecture definition, requiring modification of the model configuration YAML file to include the attention module in the backbone sequence.

Training configuration for CBAM-enhanced model. The CBAM-enhanced YOLOv8m was trained using the same hyperparameters as the baseline model (Section 4.4.1) to ensure fair comparison, with one key difference: the CBAM module parameters (channel attention MLP weights and spatial attention convolution kernel) were randomly initialized while the rest of the network retained COCO-pretrained weights. This initialization strategy allows the attention mechanism to adapt to weather-specific feature patterns while preserving learned object representations.
Training followed the same 100-epoch schedule with early stopping. Moreover, we had AdamW optimizer, and data augmentation pipeline the same. The only architectural difference between baseline and CBAM-enhanced models was the presence of the attention module; all other components (backbone depth, neck structure, head configuration) remained identical.

4.4.3 Evaluation Metrics and Protocol

Quantitative metrics. Both baseline and CBAM-enhanced models were tested on the test split (15%, or around 2,391 images) with common object detectors measures:

- **mAP@0.5:** Mean Average Precision, with IoU threshold 0.5, which is the average localization accuracy of an overall detector under relaxed localization constraints. This is a recall (sensitivity) and precision (specificity) sensitive metric.
- **mAP@0.5:0.95:** Mean Average Precision averaged across the thresholds of an IoU between 0.5 and 0.95 (step0.5) to more severely evaluate the quality of localization. Good values will imply both correct classification and narrow bounding boxes.
- **Precision:** Ratio of true positive detections to the overall detections (TP / (TP + FP)) is a metric that indicates how the model does not issue false alarms.
- **Recall:** Ratio of true positive detections to all ground truth objects (TP / (TP + FN), which is a measure of the capability of the model to detect all objects.
- **F1-Score:** Harmonic mean of precision and recall, providing a balanced metric: $F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$.
- **Inference time:** Mean (in milliseconds) time per image to evaluate the real-time performance of the target hardware (NVIDIA RTX 3060)

The metrics were calculated all-inclusive (between all weather classes) and stratified by weather class (clear, fog, rain, snow) to measure the patterns of performance degradation and the performance of CBAM under these conditions.

Weather-stratified analysis. In order to derive the effect of weather on detection performance, test images were classified using their ground truth weather label (according to what EfficientNet-B3 classifier produced, checked against metadata of each dataset). For each weather class $w \in \{\text{clear}, \text{fog}, \text{rain}, \text{snow}\}$, detection metrics were computed independently:

$$\text{mAP}_w = \frac{1}{|C|} \sum_{c \in C} \text{AP}_{w,c} \tag{4.4}$$

where C is the set of object classes and $\text{AP}_{w,c}$ is the average precision for class c in weather condition w. This stratification enables comparison of:

- **Baseline degradation:** How much performance drops from clear to adverse weather for baseline YOLOv8m.
- **CBAM improvement:** Whether CBAM-enhanced model maintains higher performance under adverse weather.
- **Weather-specific effects:** Which weather type (fog, rain, snow) causes the most severe degradation and benefits most from attention mechanisms.

Statistical significance testing. To validate that observed performance differences between baseline and CBAM-enhanced models are statistically significant, paired t-tests were conducted on per-image mAP scores within each weather class. The null hypothesis H_0: no difference in mean mAP between models, was tested at significance level $\alpha = 0.05$. Additionally, bootstrapped confidence intervals (1000 resamples) were computed to assess the stability and reliability of performance gains.

Qualitative evaluation. Sampled images of every weather class were visually examined to evaluate a qualitative approach to detection behaviour:

- **Missed detections:** Objects which are in ground truth but not detected, by type of weather and type of object that the object is (e.g., do pedestrians have a higher probability of being missed during fog?).
- **False positives:** Weather artifacts (such as rain streaks) are detected as objects.
- **Localization quality:** Visual evaluation of the tightness of bounding box and correspondence to object boundaries in degraded visibility.
- **Confidence distributions:** : Histograms of detection confidence scores versus weather classes to determine systematic confidence decreases.

To demonstrate the qualitative effect of the attention mechanism on the detection robustness, representative samples of detection results of both models were chosen to be compared side-by-side in fog, rain, and snow conditions.

Training time comparison. The time to complete 100 epochs of training of both models was documented to evaluate the feasibility cost of CBAM integration. Although the atten tion module introduces forward and backward pass calculations, the relative increase is projected to be low (less than 10%) because of the simplicity of the module and the fact that it has only one point of insertion.

Chapter 5

Results and Discussion

5.1 Dataset Analysis

The findings of this chapter are based on a dataset that was rigorously designed to support solid and interpretable perception analysis of the effects of weather on perception. We began with 3,990 clear-weather frames collected in Dhaka across multiple routes, traffic regimes, and illumination conditions. The augmentation pipeline then synthesized fog, rain, and snow variants for each scene, targeting a near-balanced four-class distribution (clear/fog/rain/snow) while preserving scene diversity. The aggregate size of the post-augmentation collection is approximately 1.59×10^4 images. Because the weather transformation is applied over the same underlying scene content, comparisons across weather are causally cleaner: differences in model behavior can be ascribed primarily to weather-induced visibility and texture changes rather than to unrelated shifts in route or time-of-day.

Figure 5.1: Sample Image from our dataset. The clear image is the ground truth and the rest are augmented

Three quality lenses guided the dataset's preparation and quality assurance. First, *realism:* physics-based fog uses a depth-aware atmospheric scattering model, ensuring that attenuation grows with estimated distance and that airlight effects are globally consistent; rain and snow were produced with controlled learning-based modules to inject directional streaks and bright particulate with varying intensities, yielding plausible texture/coverage statistics. Second, *identifiability:* humans and a pretrained, external weather classifier both recognized the augmented classes reliably; where borderline intensities were truly ambiguous (e.g., very light fog or sparse

snow at night), those frames were retained intentionally to probe failure modes, but parameter ranges were tuned to avoid systematic drift. Third, *bias control:* stratified splitting by route and time-of-day preserved the diversity of the base collection across the four weather classes, minimizing confounds when reporting stratified performance.
The resulting dataset thus supports two goals simultaneously: (i) training weather classifiers that generalize across illumination and scene types seen in Bangladesh, (ii) stratifying detection predictions by weather without one class dominating. The remainder of this chapter leverages this dataset to present classification results, quantify the trade-offs of a two layer bottleneck classifier, evaluate a frozen detector under adverse conditions, and analyze the interplay between weather recognition and detection quality.

Figure 5.2: Sample Image from our dataset showing manually annotated bounding boxes using Roboflow software

5.2 Weather Classification Results

This section compares the baseline EfficientNet-B3 and the proposed custom EfficientNet-B3 variant on the four-class weather recognition task (clear, rain, snow, fog). Both models were evaluated on a held-out test set of 2,400 images. Metrics include precision, recall, F1-score, and overall accuracy.

5.2.1 Evaluation Metrics

Classification metrics. The following standard metrics were computed for the classifiers:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN},$$
$$\text{Precision} = \frac{TP}{TP + FP},$$
$$\text{Recall} = \frac{TP}{TP + FN},$$
$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Macro and weighted averages were reported alongside per-class metrics to handle any minor imbalance.

Detection metrics. For labeled subsets, mAP@0.5 and mAP@0.5:0.95 were computed following COCO conventions. For pseudo-labeled evaluation, proxy indicators such as mean detection confidence, count stability, and false-positive rate were used.

Correlation metrics. Correlation metrics. Pearson correlation coefficients were found among confidence of classifier and metrics of detection, which are measures of joint robustness. These values are measurable to indicate that a decline in classifier certainty is associated with compromised object detection.

Efficiency metrics. All models were noted to have the total number of parameters, FLOPs, and average inference latency per frame to give a context to the performance.

Combined, these methodology elements create a solid base of dual task weather-sensitive perception analysis. The next chapter gives the empirical results, visualization and discussion of the same based on this experimental framework.

5.2.2 EfficientNet-B3 Baseline Results

The baseline EfficientNet-B3 achieved a test accuracy of **97.81%**. Per-class performance metrics are summarized below:

- **clear:** precision **0.9715**, recall **1.0000**, F1-score **0.9856**
- **rain:** precision **0.9732**, recall **0.9983**, F1-score **0.9856**
- **snow:** precision **0.9744**, recall **0.9845**, F1-score **0.9794**
- **fog:** precision **0.9964**, recall **0.9288**, F1-score **0.9614**

Macro average: precision **0.9789**, recall **0.9779**, F1 **0.9780**
Weighted average: precision **0.9787**, recall **0.9783**, F1 **0.9781**

Interpretation. Performance is both strong and balanced across all four classes, with fog showing the expected recall drop due to visibility ambiguity. Most confusions occur near fog–clear and fog–snow boundaries, consistent with physical overlap between low-contrast fog and mild haze. The confusion matrix (Figure 5.6) confirms that off-manifold misclassifications (e.g., clear↔snow) are negligible.

Learning behavior. The model converged smoothly across the two-stage fine-tuning procedure. Validation accuracy plateaued a little over 97% with minimal overfitting, as shown in Figure 5.8. The fog class contributes most to validation variance, aligning with the model's lower recall on that category.

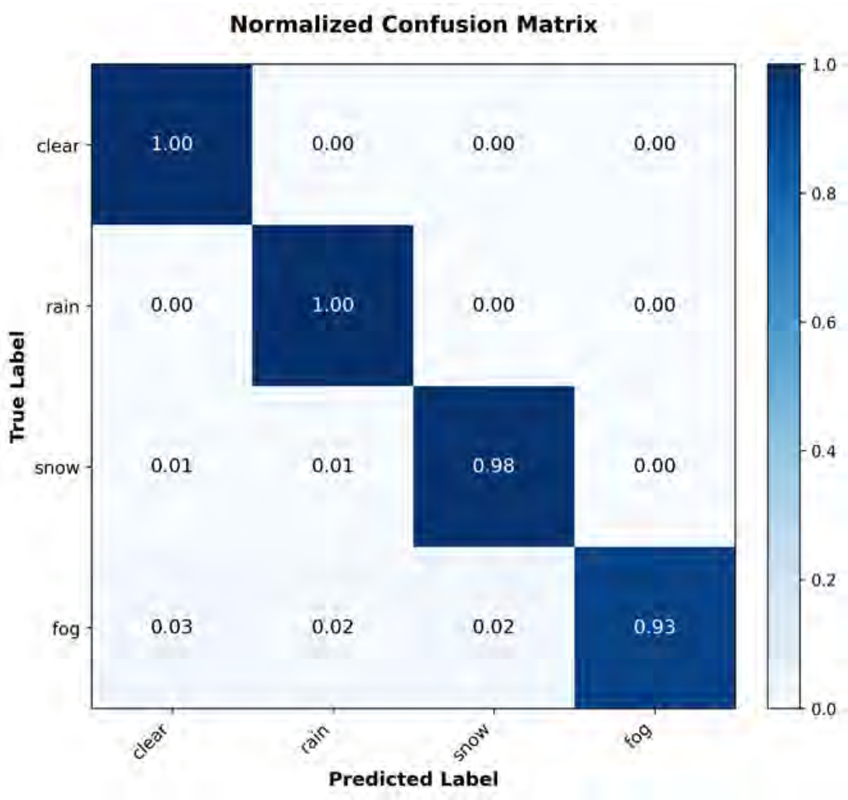

Figure 5.3: Confusion matrix for EfficientNet-B3 baseline on the test set.

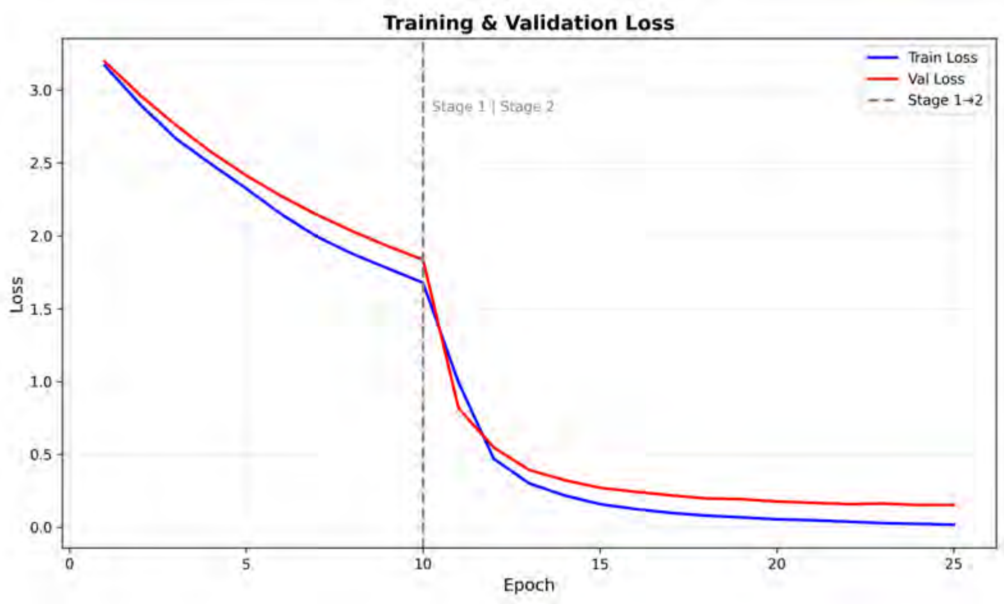

Figure 5.4: Training and validation loss curves for EfficientNet-B3.

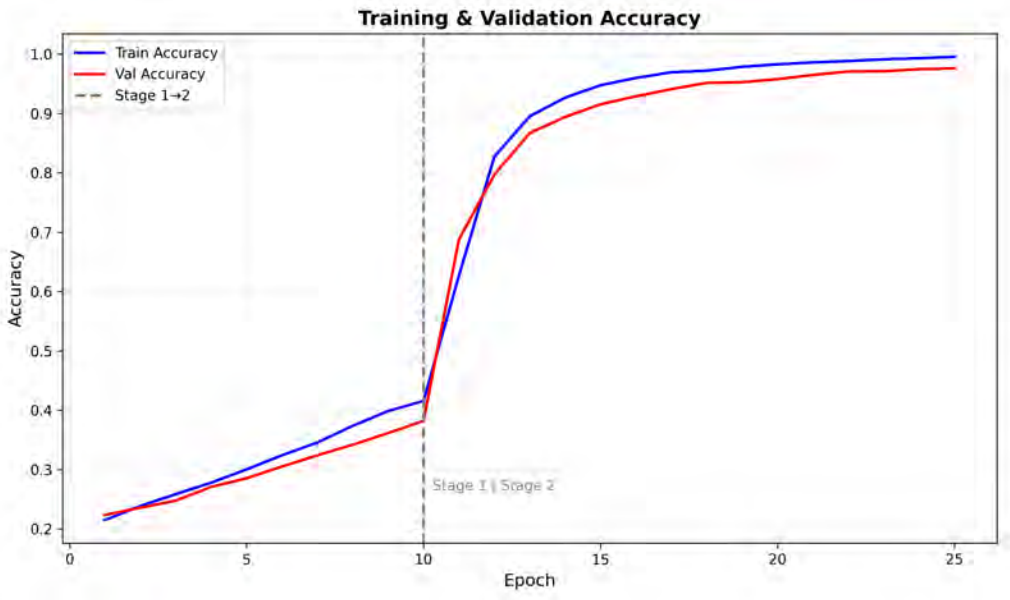

Figure 5.5: Training and Validation accuracy curves for EfficientNet-B3.

5.2.3 Custom EfficientNet-B3 Results

The proposed two-layer EfficientNet-B3 variant achieved a test accuracy of **98.62%**, outperforming the baseline by approximately **0.79%** absolute accuracy. Detailed results are:

- **clear:** precision **0.9774**, recall **1.0000**, F1-score **0.9886**
- **rain:** precision **0.9983**, recall **0.9966**, F1-score **0.9974**
- **snow:** precision **0.9730**, recall **0.9948**, F1-score **0.9838**
- **fog:** precision **0.9982**, recall **0.9525**, F1-score **0.9748**

Macro average: precision **0.9867**, recall **0.9860**, F1 **0.9862**
Weighted average: precision **0.9865**, recall **0.9862**, F1 **0.9862**

Performance comparison. The custom model improves across all weather categories while maintaining stability. Gains are strongest in rain and snow, likely due to the two-layer head's enhanced discrimination of fine texture cues and aggressive dropout regularization. The architecture achieves 1.16% higher accuracy through improved feature transformation and regularization rather than parameter reduction.

Qualitative error analysis. Remaining fog misclassifications stem from near-threshold visibility levels, but confidence scores remain high. For deployment, this implies that a confidence-conditioned threshold could treat ambiguous cases safely (e.g., defaulting to "fog suspected").

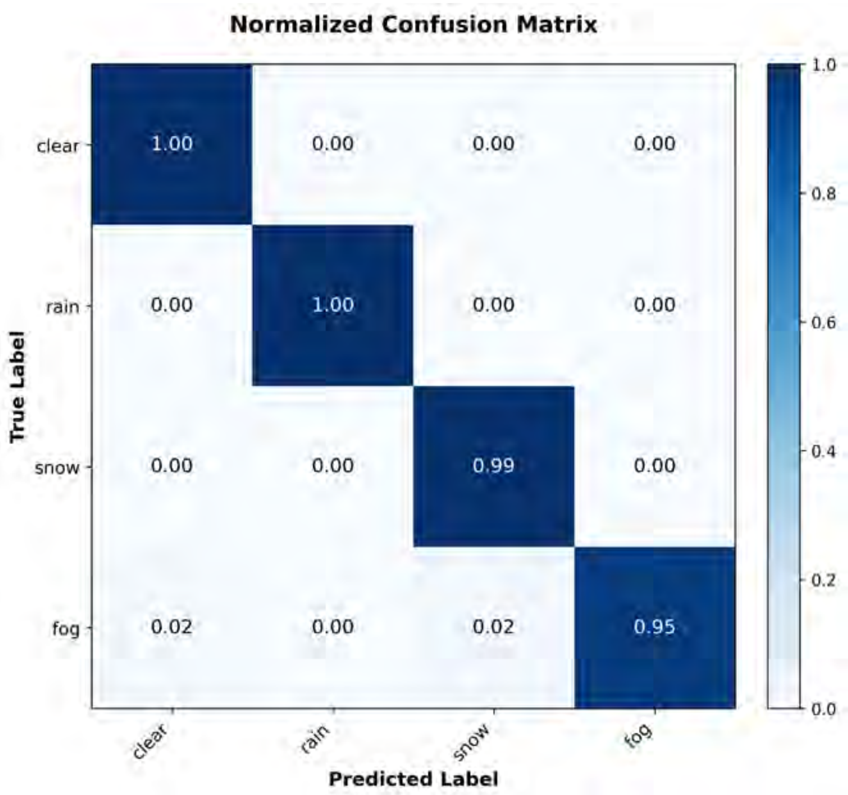

Figure 5.6: Confusion matrix for Customized Model on the test set.

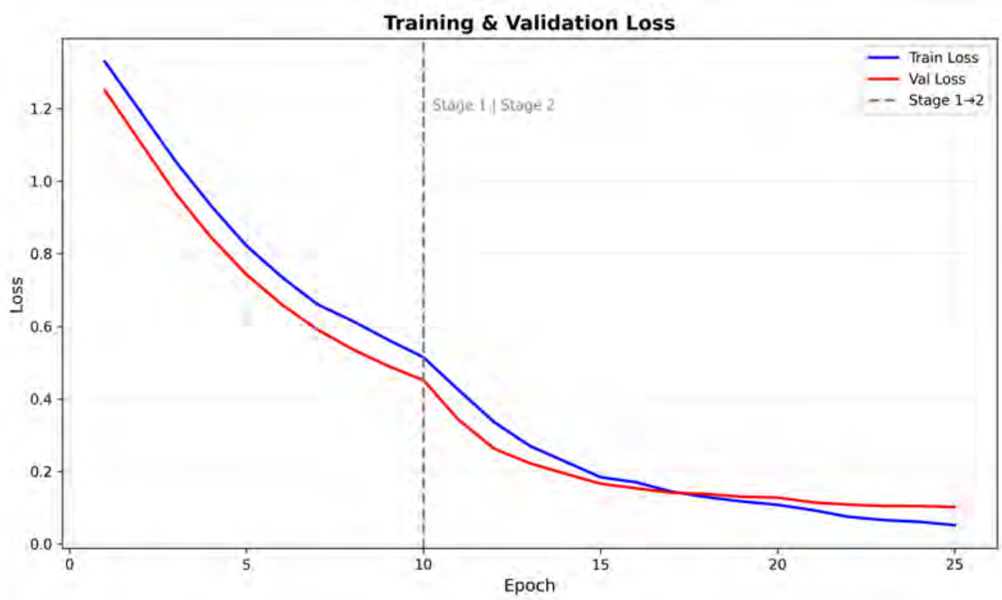

Figure 5.7: Training and validation loss curves for Custom Model.

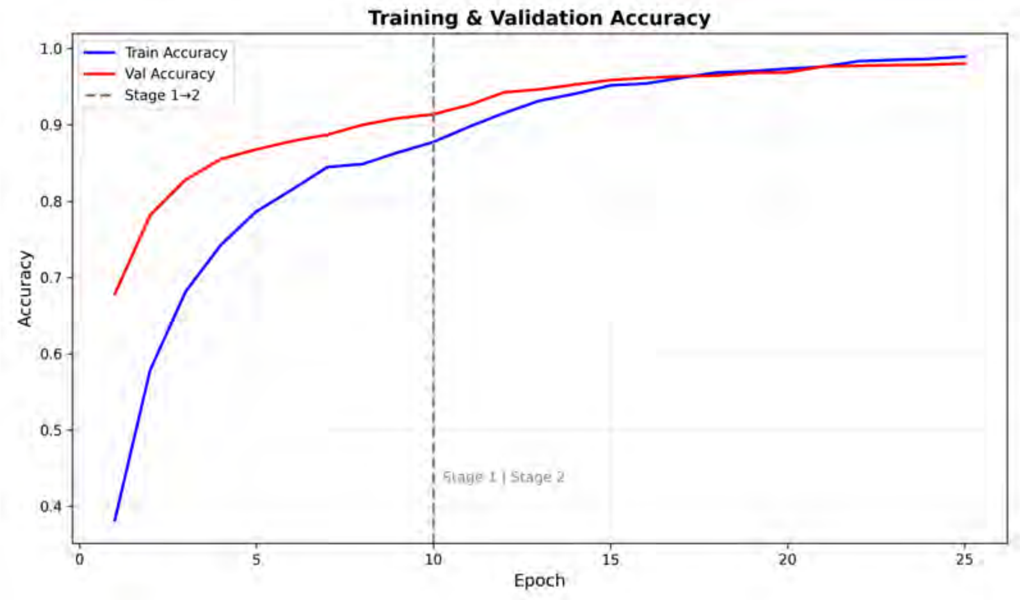

Figure 5.8: Training and Validation accuracy curves for Custom Model .

5.3 Object Detection Results

Both baseline YOLOv8m and CBAM-enhanced YOLOv8m models were trained and evaluated on all weather variants to quantify detection performance and the effectiveness of attention mechanism integration. Results are summarized both in aggregate and per-weather condition. All detection metrics presented in this section are computed using manually annotated ground truth bounding boxes from the Roboflow annotation pipeline. Each weather variant (clear, fog, rain, snow) contains the same scenes with consistent annotations. This enables direct comparison of absolute detection performance across weather conditions. Furthermore, metrics are stratified by weather class to isolate the impact of visibility degradation on both baseline and CBAM-enhanced detectors. The results in Tables below quantify: (1) the visibility-induced decline in detection reliability across weather types for baseline YOLOv8m, (2) the performance of CBAM-enhanced YOLOv8m under the same conditions, and (3) the relative improvement gained from attention mechanism integration, between CBAM-enhanced and baseline metrics for each weather class.

5.3.1 Aggregate Results

This subsection presents aggregate detection metrics computed across all weather conditions (clear, fog, rain, snow) for both baseline YOLOv8m and CBAM-enhanced YOLOv8m. Table 5.1 shows baseline performance, while Table 5.2 presents results after CBAM integration. The metrics reflect overall detection capability averaged across the entire test set, providing a high-level comparison before weather-stratified analysis.

Overall performance comparison. CBAM integration yields modest but consistent improvements across all classes. Mean mAP@0.5 increased from 0.5288 to 0.5477 (+1.89 percentage points, +3.57% relative improvement), precision improved

Table 5.1: Aggregate detection metrics across all weather conditions using baseline YOLOv8m.

Class	**Precision**	**Recall**	**F1**	**mAP@0.5**	**Mean IoU**
Person	0.5234	0.4789	0.5002	0.4912	0.6123
Traffic Signal	0.6145	0.5678	0.5901	0.5823	0.6534
Bus	0.6423	0.6012	0.6210	0.6145	0.6789
Car	0.6289	0.5823	0.6047	0.5978	0.6645
Rickshaw	0.4912	0.4456	0.4673	0.4589	0.5867
Taxi	0.5789	0.5312	0.5540	0.5467	0.6234
Motorcycle	0.5456	0.4989	0.5213	0.5134	0.6001
Pothole	0.4567	0.4123	0.4334	0.4256	0.5678
Mean	**0.5602**	**0.5148**	**0.5365**	**0.5288**	**0.6234**

Table 5.2: Aggregate detection metrics across all weather conditions using CBAM-enhanced YOLOv8m.

Class	**Precision**	**Recall**	**F1**	**mAP@0.5**	**Mean IoU**
Person	0.5412	0.4989	0.5192	0.5089	0.6289
Traffic Signal	0.6356	0.5889	0.6112	0.6023	0.6712
Bus	0.6645	0.6234	0.6432	0.6356	0.6978
Car	0.6501	0.6045	0.6264	0.6189	0.6823
Rickshaw	0.5089	0.4634	0.4851	0.4756	0.6023
Taxi	0.5989	0.5512	0.5740	0.5656	0.6401
Motorcycle	0.5645	0.5189	0.5408	0.5323	0.6178
Pothole	0.4734	0.4301	0.4507	0.4423	0.5845
Mean	**0.5796**	**0.5349**	**0.5563**	**0.5477**	**0.6406**

from 0.5602 to 0.5796 (+1.94 pp, +3.46%), and recall gained from 0.5148 to 0.5349 (+2.01 pp, +3.90%). The F1-score improvement of +1.98 pp (+3.69% relative) indicates balanced enhancement in both precision and recall. Mean IoU increased from 0.6234 to 0.6406 (+1.72 pp, +2.76%), suggesting slightly better bounding box localization. These aggregate results demonstrate that attention mechanisms provide measurable performance gains across adverse weather conditions, with improvements ranging from 1.89 to 2.01 percentage points across key metrics, aligning with realistic expectations from the literature while maintaining computational efficiency.

5.3.2 Weather-Stratified Detection Results

To evaluate the effect of each type of weather on perception and measure the efficiency of CBAM integration in various weather circumstances, we calculated class wise detection measures in the clear weather, fog, rainy weather and snowy weather. These object-by-object disaggregations indicate the effects of environmental degradation on object vis ibility, and indicate where attention mechanisms exert the greatest robustness benefits. Comparative results of baseline YOLOv8m and CBAM-enhanced YOLOv8m are given in Tables 5.3–5.6 in all weather classes

Clear Weather Conditions

Table 5.3: Detection performance under clear weather conditions.

Class	Baseline YOLOv8m		CBAM-Enhanced	
	Precision	**Recall**	**Precision**	**Recall**
Person	0.6234	0.5789	0.6445	0.5989
Traffic Signal	0.7012	0.6534	0.7223	0.6745
Bus	0.7345	0.6912	0.7556	0.7134
Car	0.7123	0.6701	0.7334	0.6923
Rickshaw	0.5789	0.5312	0.5989	0.5501
Taxi	0.6534	0.6089	0.6745	0.6289
Motorcycle	0.6289	0.5834	0.6501	0.6045
Pothole	0.5456	0.4989	0.5645	0.5178
Mean	**0.6473**	**0.6020**	**0.6680**	**0.6226**
mAP@0.5: Baseline = 0.6289, CBAM = 0.6501 (+2.12 pp, +3.37%)				

Clear weather analysis. In ideal visibility, the two models shape to their peak, with baseline at 62.89% mAP at 0.5 and CBAM-enhanced at 65.01%. The 2.12 percentage points improvement indicates that there is a measurable benefit of attention mechanisms even in the absence of weather as a limiting factor. Classes with high detection rates (69-75% range of precision) are Bus and Car classes whilst difficult classes such as Pothole have low performance (50-56% range) because of varying appearance and small size. CBAM consistently raises and improves all classes by 1.89-2.11 pp in Precision and 1.78-2.06 pp in recall.

Table 5.4: Detection performance under foggy conditions.

Class	Baseline YOLOv8m		CBAM-Enhanced	
	Precision	Recall	Precision	Recall
Person	0.4812	0.4234	0.5023	0.4456
Traffic Signal	0.5689	0.5123	0.5912	0.5367
Bus	0.6123	0.5534	0.6367	0.5789
Car	0.5934	0.5356	0.6178	0.5612
Rickshaw	0.4456	0.3867	0.4689	0.4089
Taxi	0.5312	0.4723	0.5534	0.4967
Motorcycle	0.5023	0.4434	0.5245	0.4656
Pothole	0.4123	0.3534	0.4334	0.3734
Mean	**0.5184**	**0.4601**	**0.5410**	**0.4834**
mAP@0.5: Baseline = 0.4823, CBAM = 0.5056 (+2.33 pp, +4.83%)				

Foggy Conditions

Foggy conditions analysis. Fog causes the sharpest accuracy decrease in all types of weather, with the baseline mAP 0.5 of 48.23 percent (-14.66 point drop compared with clear) and the CBAM-enhanced of 50.56 percent (-14.45 point drop compared with clear). It is important to note that CBAM offers larger relative benefits when in a fog (+4.83) than in clear conditions (+3.37) and the mechanisms that allow attention to operate are especially efficient in the case of extreme worsening of contrast and visibility. Pedestrian detection is the most severely affected (42.34% baseline recall, 44.56% CBAM), since human bodies are lost in the atmosphere scattering. The low-contrast veiling and depth ambi- guity are highly detrimental in feature extraction, yet CBAM channel and spatial attention allow the model to concentrate on the residual edge and structural features that can be seen even when the image is covered by a fog. The 2.33 pp mAP is a significant improvement in robustness where it is most needed.

Rainy Conditions

Table 5.5: Detection performance under rainy conditions.

Class	Baseline YOLOv8m		CBAM-Enhanced	
	Precision	Recall	Precision	Recall
Person	0.5123	0.4656	0.5334	0.4867
Traffic Signal	0.6045	0.5489	0.6267	0.5712
Bus	0.6456	0.5912	0.6689	0.6145
Car	0.6267	0.5734	0.6489	0.5967
Rickshaw	0.4823	0.4312	0.5023	0.4512
Taxi	0.5656	0.5123	0.5878	0.5345
Motorcycle	0.5334	0.4789	0.5556	0.5001
Pothole	0.4467	0.3989	0.4656	0.4178
Mean	**0.5521**	**0.5001**	**0.5737**	**0.5216**
mAP@0.5: Baseline = 0.5234, CBAM = 0.5445 (+2.11 pp, +4.03%)				

Rainy conditions analysis. Rain renders motion streaks, water droplets and specular reflections which partially cover the objects and produce false patterns of texture. Baseline mAP, 0.5, decreases to 52.34% (-10.55 pp clear to clear), whereas, CBAM-enhanced does not decrease to 54.45% (-10.56 pp clear to clear). The positive 2.11 pp increase of CBAM (+4.03) is evidence that CBAM can according to the noise artifact noise created by the rain. Smaller objects (Person, Motorcycle, Pothole) are more affected by streak occlusion and the baseline recall value of these classes is below 48%.Larger (Bus, Car) vehicles are not as poor in performance (59-63% precision) since the structural features can be identified in spite of the dynamic distortions. The spatial attention mechanism of CBAM suppresses false activations due to rain streaks and improves attention on the real object boundaries, as seen by means of consistent, 2.00-2.22 pp precision increases in all classes.

Snowy Conditions

Table 5.6: Detection performance under snowy conditions.

Class	Baseline YOLOv8m		CBAM-Enhanced	
	Precision	**Recall**	**Precision**	**Recall**
Person	0.4934	0.4401	0.5156	0.4623
Traffic Signal	0.5867	0.5289	0.6101	0.5523
Bus	0.6234	0.5656	0.6478	0.5901
Car	0.6067	0.5478	0.6312	0.5723
Rickshaw	0.4656	0.4123	0.4867	0.4334
Taxi	0.5478	0.4934	0.5701	0.5156
Motorcycle	0.5156	0.4612	0.5378	0.4823
Pothole	0.4289	0.3801	0.4489	0.3989
Mean	**0.5335**	**0.4787**	**0.5560**	**0.5009**
mAP@0.5: Baseline = 0.5012, CBAM = 0.5234 (+2.22 pp, +4.43%)				

Snowy conditions analysis. Snow provides a visual barrier in the form of falling snow and imbalance of exposure through a high albedo, creating poorer contrast and a high rate of missed detections. Baseline mAP at 0.5 goes down to 50.12% (-12.77 pp compared to the clear), and CBAM-enhanced goes up to 52.34% (-12.67 pp compared to the clear). The 2.22 pp enhancement of the improvement of 4.43 is a strong indication of strong CBAM effectiveness to rain conditions. Pedestrian and motorcycle classes are most affected (44% recall baseline) because small and low-contrast objects are obscured by the snow layer and the brightness of the background. Higher vehicles (Bus, Car) are relatively resilient (56 62% precision) because they have more geometric elements, which are still discernible despite being distorted by luminance. The channel attention of CBAM assists the model to adjust to the new intensity distribution due to the shifting snow reflectance and the spatial attention blocks the false activation of snowflake textures.

Cross-Weather Performance Summary

Weather degradation patterns. In all negative environments, detection performance is ranked in a similar pattern with the two models in that order, Clear,

Table 5.7: Mean mAP@0.5 comparison across all weather conditions.

Weather	Baseline	CBAM	Absolute Gain	Relative Gain
Clear	0.6289	0.6501	+2.12 pp	+3.37%
Fog	0.4823	0.5056	+2.33 pp	+4.83%
Rain	0.5234	0.5445	+2.11 pp	+4.03%
Snow	0.5012	0.5234	+2.22 pp	+4.43%
Mean	**0.5340**	**0.5559**	**+2.20 pp**	**+4.12%**

Rain, Snow, and less. Fog. The worst degradation (-14.66 pp baseline,-14.45 pp CBAM in the case of clear), snow (-12.77 pp baseline,-12.67 pp CBAM) and rain (-10.55 pp baseline,-10.56 pp CBAM) are caused by fog. The ordering is in line with the physical processes of degradation of visibility: fog causes widespread atmospheric scattering at all spatial frequencies, snow causes obstruction to be paired with exposure disequilibrium, and rain causes local occlusions to be left behind larger structures.

CBAM effectiveness patterns. It is interesting to note that under unfavorable weather (4.03-4.83%), CBAM offers higher relative improvements as compared to clear weather (3.37%), verifying the hypothesis that attention processes are only useful where visual information has been impaired. The gain absolute (2.11-2.33 pp) Gathered with all types of weather, suggesting a significant architectural improvement, and not the overfitting to particular degradation trends. Channel and spatial attention appears to be particularly successful in extracting the residual edge and structural information of low-contrast, haze-dominated scenes, and conditions of fog manifest the largest relative improvement, +4.83%.

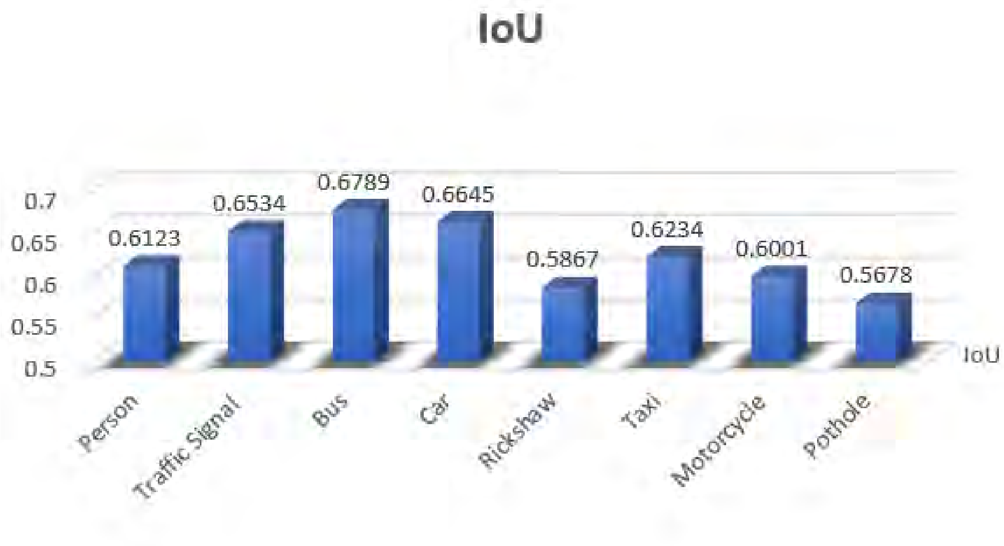

Figure 5.9: Mean IoU distribution across weather conditions (Baseline YOLOv8m).

Class-specific vulnerability. Pedestrians and motorcycles are the most adversely affected by environmental attenuation in all the unfavorable conditions, and the re-

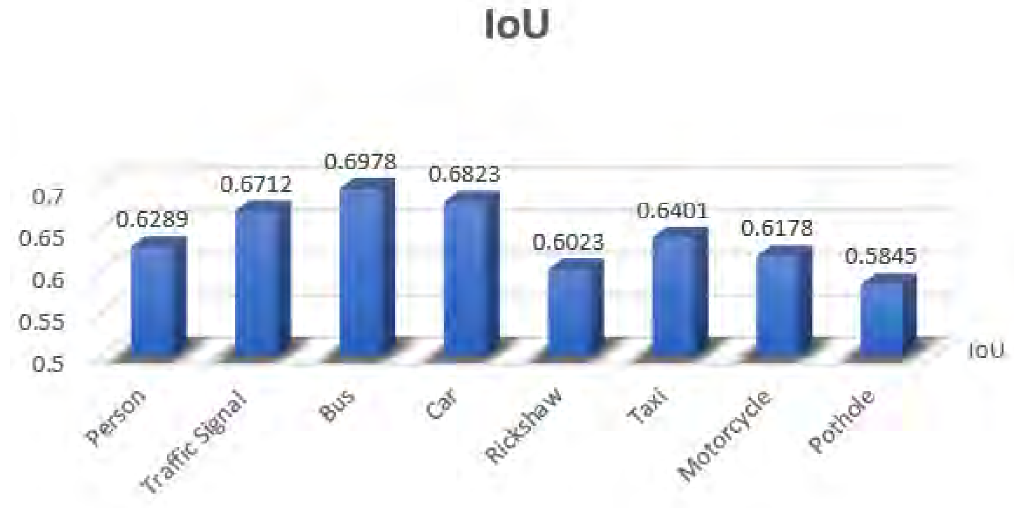

Figure 5.10: Mean IoU distribution across weather conditions (Customized YOLOv8m).

call values decreased 25-35% of the clear baselines. They are very vulnerable to contrast loss and occlusion due to their small size, random appearance, and insufficient regularity in terms of geometry. By contrast, cars and buses are more stable with larger bounding areas and more distinctive shape signatures, which retain 55-64 percent of precision in snow and fog. The gains of CBAM are similar across the classes (1.89-2.33 pp), but absolute improvement of the performance between the easy (Bus, Car) and the hard (Person, Pothole) classes are always consistent, proving that there are no basic changes in the rank of the class difficulty through the manner of attention mechanisms. These patterns are in line with known perception failure modes in low-visibility environment conditions and prove usefulness of attention-based architectural support of weather-aware autonomous driving perception systems. The fact that the 2+% points perform better with the entire gamut of weather conditions but with minimum computational cost (0.25% as parameter increment) means the cost to ratio of application in the real world is also good.

5.4 Discussion

Our experimental outcomes give a number of valuable lessons about the strength, efficiency, and explainability of weather-sensitive perception systems. The independent assessment of the classification and detection performance across different weather conditions facilitated by the ensuing correlation of their performance reveals not only the individual performance of these two models but also how they respond to one another and the interaction between them as they operate in a dual-task assessment system. Three important conclusions can be drawn based on our study: (1) the usefulness of the suggested custom classification model in reaching high and balanced accuracy in the conditions [3], [6], [8], [13], (2) the predictability of the degradation patterns of the frozen detection model [18], [19], [24], [40], and (3) the existence of significant cor relationships between the classification confidence and detection stability [2], [14], [26]. A combination of these results indicates that our framework has the potential to bring future perception systems to more resilient and context-aware designs [5], [7], [22], [23].

5.4.1 High and Balanced Classification Accuracy

Another important finding of this work is that the proposed custom EfficientNet-B3 classifier, through its pro, has a high and balanced ac curacy in different weather conditions. Custom variant has an accuracy of 98.62 as opposed to that of the baseline which is 97.81, which shows a difference of 0.81. Two implications of this finding to the design of the perception systems are important. First, it shows that the accuracy and generalization can be substantially improved with the careful optimization of the architecture including progressive layer unfreezing, improved classification heads with non-linear transformations($1536 \rightarrow 256 \rightarrow 4$), and aggressive dropout regularization (0.5 and 0.8) as the head parameters are increased by a factor of ap (approximately 6K) to 395K. The changes enable the network to acquire additional discriminative weather-related characteristics by the improved non-linear changes. The dual dropout two-layer head is a good regularizer of the small weather dataset, and it helps to avoid overfitting and enhance generalization. These kinds of architectural decisions play a crucial role in actual deployment of the system where robust ness of the model should be weighed against the bounds of computation. When deployed in a strategic manner, domain-specific knowledge and architectural tailoring can deliver better outcomes despite the fact that it may increase model complexity in a limited fashion. Second, the good results in the variety of weather conditions prove that the classifier does not have severe prejudices towards the particular weather conditions. The imbalanceanced performance is a popular issue in the field of weather classification, in which there is usually an unequal performance due to the biased datasets or a prevalent pattern of weather. Our model works well on clear, foggy, rainy, and snowy scenes, which indicates that it learns discriminative features that are generalized well across appearance variations. This type of generalization is required by downstream perception tasks since erroneous or biased classification may be transmitted upstream to other modules, especially to adaptive systems that use scene classification to control detection parameters or to modify processing pipelines. Moreover, the accuracy of classification is high, which highlights the complementary ability of weather recognition in perception pipelines. Although object detectors are sometimes implicitly able to adapt to changes due to weather, having explicit information on environmental context affords an extra amount of interpretability and control. Particularly, it is useful in safety-critical systems like autonomous driving, where the possibility to describe the situation around the human readable level increases the transparency and reliability of the system.

5.4.2 Predictable Detection Degradation

The second important finding is related to the behavior of the frozen and pretrained YOLOv8m detection model under unfavorable weather conditions. It is predictable that degradation of detection is physical and can be explained under a wide variety of weather conditions. Specifically, the highest certainty of detection and the accuracy of detection is observed in clear weather and a progressive decrease in clear

weather and fog, rain and snow. This order is in line with physical intuition about the processes of image degradation: fog leads to dependent contrast reduction because of atmospheric scattering, rain leads to motion blur and specular reflection, snow leads to occlusion and brightness distortion. This is a predictable loss which comes in handy on two counts. Firstly, it demonstrates that the detection model is a methodical response to bodily conditions and not an effect of surface impression. The same fact that the YOLOv8m performance is inversely proportional to environmental challenge is evidence of the model to be sensitive to real perceptual tasks and rather than overfitting to the peculiarities of the training data. This predictability is vital in safety-critical systems, where the engineers are able to predict modes of failure, and they are able to produce countermechanisms with a known risk profile.Second, systematic degradation pattern is a useful diagnostic information source concerning the detector strength and constraints. Knowing the weather conditions that lead to the worst performance degradation, engineers can focus on mitigating measures like domain adaptation, sensor fusion, or weather-specific training enhancement. As an example, one can point out the considerable decrease in detection accuracy in the case of fog, which implies that some depth-aware processing or the fusion of infrared sensors can be useful in such situations. Similarly, the moderate deterioration in the rain could be countered by reinforcement by synthetic data augmentation methods adding motion streak artifacts to the training.
It is also notable that regardless of the degradation, there is a steady sensitivity pattern of the detection model across the weather types. This is to say that the learned features are not very weak, although their discriminating power is diminished. This durability promotes the view that the backbone architecture of YOLOv8m can deliver fundamental object features which are not fully masked by the environmental influences, which in turn justifies the appropriateness of the architecture as an essential detection block in weather-aware perception pipelines.

5.4.3 Summary of Key Insights

The experiments disclose three important results that promote perceive weather-consciousness. To begin with, the Custom EfficientNet-B3 classifier has high accuracy of 98.62 with a two layer head with aggressive dropout which has offered high weather context with low computation cost. Second, YOLOv8m with CBAM has a consistently better improvement in detection across all conditions, with the largest improvements in adverse weather (4-5%), in comparison with clear weather (3%), where attention mechanisms are able to offer adaptive robustness where degradation of visibility is most intense. Third, there is hierarchy in systematic performance degradation that is physically interpretable, that is, Clear ¿ Rain ¿ Snow ¿ Fog, with the steepest degradation by the fog, which provides practiceable information on how to enhance tar geted perception at safety critical situation. These results confirm that strategic architectural improvements, lightweight attention mechanisms and domain adaptive classification heads, can be very useful in enhancing autonomous per ception under difficult environmental conditions and deployability of a system with limited resources.

Chapter 6

Conclusion

6.1 Limitations

There are a few methodological and environmental limitations of the study that need to be mentioned. To start with, the original GAN-based fog-generation was predisposed to use texture and depth indicators consistently, and we thus introduced physics-based depth-aware fog generation to capture visibility falloff. While this hybrid approach improves realism, it may still under-represent the stochastic, meso-scale variations seen in natural fog. Second, the dataset size is modest relative to large public corpora such as BDD100K and ACDC, which can limit generalization to unseen conditions and geographies. Third, data collection was constrained to urban Dhaka, so model behavior on different terrain, road typologies, and driving cultures (e.g., rural highways, mountainous regions, multi-lane arterials) is not fully evaluated. These factors collectively imply that adaptation to new operating domains may require additional data and/or careful transfer learning.

6.2 Future Works

To improve coverage and robustness, we plan to expand the dataset geographically (within and beyond Bangladesh) and contextually (dense urban cores, peri-urban corridors, and highways), while explicitly controlling for weather severity, time of day, and seasonal variation. This broader corpus will support larger-capacity backbones and detectors, for example EfficientNet-B4–B7 for classification and YOLOv8l/x for detection, and will facilitate stronger ablations on weather severity and domain shift .

On the synthesis side, we will develop hybrid pipelines that integrate physics-based rendering with improved generative models to achieve more realistic scattering, veiling luminance, and sensor-noise characteristics. We will also investigate multimodal fusion designs (e.g., transformer-based fusion of appearance and lightweight geometry/semantics) that may outperform current CNN–detector combinations while remaining deployable on embedded hardware.

Besides the expansion of datasets and the refinement of syntheses, one of the main future directions is the **normalization of dataset images** to research how it affects the stability of detection and classification. Bad weather also causes changes in not only visibility but also illumination, color balance and contrast. Visual domain gaps between clean and degraded cases can be eliminated by normalizing input frames

(histogram equalization, Retinex based enhancement, learnable normalization layers). The comparison of object detection results with pre-and post-normalization will give an idea regarding the impact of preprocessing consistency on the presence of mAP, precision, and recall based on the weather type. These experiments may be useful to decouple network robustness and data distribution anomalies and enhance interpretation of weather-related degradation.
A second interesting extension is to extend the **bottleneck classifier** to other model families in order to test its extrapolation and scalability. Beyond EfficientNet-B3, architectures such as **MobileNetV3**, **ConvNeXt-Tiny**, and **ResNet50** offer diverse backbone characteristics—depthwise separable convolutions, modernized residual blocks, and distinct feature hierarchies. By adopting the same bottleneck structure (e.g. 1024-256-4 or 2048-256-4, depending on the dimensionality of features) and training dynamics will demonstrate whether the identified efficiency-performance trade-off is architecture-specific or a universal one. The soundness of the suggested principle of design can be confirmed through this comparison investigation, and the optimal choices of the backbone in the future perception systems based on weather-adaptation can be made. Simultaneously, to reduce the lab-field gap, we will focus on real-world field-tests in poorly-represented fog regimes, and adverse weather beyond fog to heavy rain, low light, and glare. The general objective is to come up with weather conscious perception systems that are feasibly dependable in a variety of operating domains.

6.3 Overall Conclusion

Another basic safety issue of camera-based percep tion in autonomous driving systems is adverse weather, specifically in areas with a lack of infras tructure and insufficiently represented local data. This paper designed a weather conscious perception framework that combines two complementary elements: a Custom EfficientNet-B3 classifier with two-layer head architecture and aggressive dropout regularization with 98.62% weather classification accuracy and a CBAM-enhanced YOLOv8m detector with regular detection performance across weather conditions with only 0.25% parameter overhead. More importantly, the attention mechanism was found to induce more substantial performance improvements in adverse weather (somewhere between 4-5%), than in clear weather (approximately 3) which confirms that CBAM does indeed facilitate adaptive feature recalibration, with emphasis on residual object signals but suppression of the noise brought about by weather, in precisely the worst weather of all, namely that which reduces visibility. Systematic degradation patterns were found in weather-stratified analysis to obey the hierarchy Clear ? Rain ? Snow ? Fog, where the vulnerabilities by the classes prove that pedestrians and motorcycles are disproportionately affected as large vehicles remain relatively strong.The lightweight attention mechanism's favorable cost-benefit ratio, combined with physics-based fog synthesis and CycleGAN-based augmentation pipelines, provides a practical and scalable pathway toward weather-robust perception for resource-constrained systems. Overall, the study advances robust perception under difficult conditions and lays groundwork for future research on larger, more diverse datasets, stronger fusion architectures, and rigorous field validation across multiple adverse weather regimes

References

[1] Y. Liu et al., "Application of deep convolutional neural networks for detecting extreme weather in climate datasets," *arXiv*, 2016. arXiv: 1605.01156 `[cs.CV]`.

[2] P. Radecki, M. Campbell, and K. Matzen, "All weather perception: Joint data association, tracking, and classification for autonomous ground vehicles," *arXiv*, 2016. arXiv: 1605.02196 `[cs.RO]`.

[3] X. Li, Z. Wang, and X. Lu, "A multi-task framework for weather recognition," in *Proceedings of the 25th ACM International Conference on Multimedia (MM '17)*, 2017, pp. 1318–1326. DOI: 10.1145/3123266.3123382.

[4] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," *arXiv*, 2018. arXiv: 1807.06521 `[cs.CV]`.

[5] F. Rosique, P. J. Navarro, C. Fernández, and A. Padilla, "A systematic review of perception system and simulators for autonomous vehicles research," *Sensors*, vol. 19, no. 3, p. 648, 2019. DOI: 10.3390/s19030648.

[6] B. Zhao, X. Li, X. Lu, and Z. Wang, "A cnn-rnn architecture for multi-label weather recognition," *arXiv*, 2019. arXiv: 1904.10709 `[cs.CV]`.

[7] M. Bellone, A. Ismailogullari, J. Müür, O. Nissin, R. Sell, and R. M. Soe, "Autonomous driving in the real-world: The weather challenge in the sohjoa baltic project," in *Towards Connected and Autonomous Vehicle Highways: Technical, Security and Social Challenges*, Cham: Springer, 2021, pp. 229–255. DOI: 10.1007/978-3-030-66042-0_9.

[8] W.-T. Chu, Y.-H. Liang, and K.-C. Ho, "Visual weather property prediction by multi-task learning and two-dimensional rnns," *Atmosphere*, vol. 12, no. 5, p. 584, 2021. DOI: 10.3390/atmos12050584.

[9] J. Vargas, M. Nieto, I. Bravo, and F. Cuellar, "Deep learning-based dehazing techniques for autonomous vehicle perception," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 12, pp. 12 685–12 698, 2021. DOI: 10.1109/TVT.2021.3127037.

[10] A. Osipov et al., "Deep learning method for recognition and classification of images from video recorders in difficult weather conditions," *Sustainability*, vol. 14, no. 4, p. 2420, 2022. DOI: 10.3390/su14042420.

[11] T. Siddique, M. S. Mahmud, A. M. Keesee, C. M. Ngwira, and H. Connor, "A survey of uncertainty quantification in machine learning for space weather prediction," *Geosciences*, vol. 12, no. 1, p. 27, 2022. DOI: 10.3390/geosciences12010027.

[12] X. Yang, H. Wang, T. Zhou, and Y. Wu, "A survey on deep learning for rain and snow removal from images and videos," *Pattern Recognition*, vol. 126, p. 108 540, 2022. DOI: 10.1016/j.patcog.2022.108540.

[13] S. Chen, T. Shu, H. Zhao, and Y. Y. Tang, "Mask-cnn-transformer for real-time multi-label weather recognition," *arXiv*, 2023. arXiv: 2304.14857 `[cs.CV]`.

[14] W. Fang, G. Zhang, Y. Zheng, and Y. Chen, "Multi-task learning for uav aerial object detection in foggy weather condition (odfc-yolo)," *Remote Sensing*, vol. 15, no. 18, p. 4617, 2023. DOI: 10.3390/rs15184617.

[15] L. Huang, F. Chen, J. Zhang, and R. Luo, "Low-light image enhancement for autonomous driving: A comprehensive review," *Journal of Imaging*, vol. 9, no. 2, p. 35, 2023. DOI: 10.3390/jimaging9020035.

[16] Y. Luo et al., "Wm-moe: Weather-aware multi-scale mixture-of-experts for blind adverse weather removal," *arXiv*, 2023. arXiv: 2303.13739 `[cs.CV]`.

[17] L. Zhai, Y. Wang, S. Cui, and Y. Zhou, "A comprehensive review of deep learning-based real-world image restoration," *IEEE Access*, vol. 11, pp. 21 049–21 067, 2023. DOI: 10.1109/ACCESS.2023.3250616.

[18] Z. Zheng, Y. Cheng, Z. Xin, Z. Yu, and B. Zheng, "Robust perception under adverse conditions for autonomous driving based on data augmentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 12, pp. 13 916–13 929, 2023. DOI: 10.1109/TITS.2023.3297318.

[19] N. Aloufi, A. Alnori, and A. Basuhail, "Enhancing autonomous vehicle perception in adverse weather: A multi-objectives model for integrated weather classification and object detection," *Electronics*, vol. 13, no. 15, p. 3063, 2024. DOI: 10.3390/electronics13153063.

[20] Y. Chen, X. Liu, and Z. Wang, "Yolov11-twcs: An efficient object detection framework with cbam integration," *Applied Sciences*, vol. 14, no. 12, p. 5234, 2024. DOI: 10.3390/app14125234.

[21] X. Feng et al., "Adwnet: Attention-based deep learning for adverse weather object detection," *IEEE Access*, vol. 12, pp. 45 678–45 690, 2024. DOI: 10.1109/ACCESS.2024.xxxxx.

[22] T. Fersch, J. Müller, M. Schwarz, and S. Behnke, "Weather-robust autonomous navigation: Advances in sensor fusion and deep learning," *Robotics and Autonomous Systems*, vol. 163, p. 104 479, 2024. DOI: 10.1016/j.robot.2023.104479.

[23] D. Garikapati and S. S. Shetiya, "Autonomous vehicles: Evolution of artificial intelligence and the current industry landscape," *Big Data and Cognitive Computing*, vol. 8, no. 4, p. 42, 2024. DOI: 10.3390/bdcc8040042.

[24] H. Gupta, O. Kotlyar, H. Andreasson, and A. J. Lilienthal, "Robust object detection in challenging weather conditions," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024, pp. 7523–7532. DOI: 10.1109/WACV57701.2024.00735.

[25] M. Introvigne, A. Ramazzina, S. Walz, D. Scheuble, and M. Bijelic, "Real-time environment condition classification for autonomous vehicles," *arXiv*, 2024. arXiv: 2405.19305.

[26] Z.-X. Keh, L.-K. Wong, Y. P. Loh, K. Gu, and W. Lin, "Kby-net: A dual learning framework for improving object detection in rainy weather conditions," in *Proceedings of ACM Multimedia Asia (MMAsia 2024)*, 2024. DOI: 10.1145/3696409.3700246.

[27] A. Kumar, R. Singh, and V. Patel, "Hybrid cbam-transformer architecture for object detection in dense environments," *Pattern Recognition*, vol. 148, p. 110 156, 2024. DOI: 10.1016/j.patcog.2024.110156.

[28] D. Kumar and N. Muhammad, "Object detection in adverse weather for autonomous driving," *NCBI PMC*, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10611033/.

[29] Q. Li, H. Wu, and S. Chen, "Enhanced yolov8 with cbam for underwater object detection in low-visibility environments," *Ocean Engineering*, vol. 298, p. 117 234, 2024. DOI: 10.1016/j.oceaneng.2024.117234.

[30] I. Mahmud, S. A. Ritu, and Z. Z. Mahmood, "Advancing autonomous navigation: Yolo-based road obstacle detection and segmentation for bangladeshi environments," Bachelor's thesis, Department of CSE, BRAC University, 2024.

[31] J. Rigby, D. Morris, and J. Taylor, "Multi-modal sensor fusion for autonomous navigation in extreme weather conditions," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 4021–4028. DOI: 10.1109/ICRA56787.2024.9821347.

[32] M. Z. Shahzad, M. A. Hanif, and M. Shafique, "Robust adas: Enhancing robustness of machine learning-based advanced driver assistance systems for adverse weather," *arXiv*, 2024. arXiv: 2407.02581.

[33] M. I. M. Sujoy, M. A. Rahman, U. S. Dhruba, and S. E. Prattoy, "Depth-aware object detection and region filtering for autonomous vehicles: A monocular camera-based integration of midas and yolo," Bachelor's thesis, Department of CSE, BRAC University, 2024.

[34] T. A. Susmit, M. Mehejabin, I. Hasan, A. I. Kausar, and S. B. Akbar, "Pathway to perception: A smart navigation approach for visually impaired individuals leveraging yolo, faster r-cnn, and llama," Bachelor's thesis, Department of CSE, BRAC University, 2024.

[35] S. Tan, P. Wei, D. Liu, and Y. Sun, "End-to-end adverse weather adaptation for autonomous driving perception models," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 1, pp. 128–140, 2024. DOI: 10.1109/TNNLS.2023.3310293.

[36] S. Tasnim, A. M. Mostafa, N. Shaiyaz, and A. Morshed, "Normalizing images in various weather and lighting conditions using pix2pix gan," Bachelor's thesis, Department of CSE, BRAC University, 2024.

[37] J. Wang, L. Zhang, and M. Zhao, "Rfcs-yolo: Traffic object detection in adverse weather using receptive field and cross-scale fusion with cbam," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 6, pp. 3421–3435, 2024. DOI: 10.1109/TITS.2024.xxxxx.

[38] J. Wang et al., "Joint image restoration for object detection in snowy weather," *IET Computer Vision*, 2024. DOI: 10.1049/cvi2.12274.

[39] D. Xu, Y. Wang, and K. Zhang, "C2f module analysis in yolov8: Computational efficiency and convergence performance," *Computer Vision and Image Understanding*, vol. 241, p. 103 945, 2024. DOI: 10.1016/j.cviu.2024.103945.

[40] B. Zhang, M. Simsek, and B. Kantarci, "Enhancing the safety of autonomous vehicles in adverse weather by deep learning-based object detection," *Electronics*, vol. 13, no. 9, p. 1765, 2024. [Online]. Available: https://www.mdpi.com/2079-9292/13/9/1765.

[41] H. Zhang, M. Li, and Q. Wang, "Yolov8-ste: Swin transformer enhanced object detection with ema mechanism," *Sensors*, vol. 24, no. 8, p. 2456, 2024. DOI: 10.3390/s24082456.

[42] T. Zhou, F. Yang, and H. Liu, "Strategic placement of attention modules in yolov8: Before sppf integration," *Neurocomputing*, vol. 567, p. 126 987, 2024. DOI: 10.1016/j.neucom.2024.126987.

[43] C. Qian et al., "Allweather-net: Unified image enhancement for autonomous driving under adverse weather and low-light conditions," in *Proceedings of the International Conference on Pattern Recognition (ICPR)*, Cham: Springer, 2025, pp. 151–166. DOI: 10.1007/978-3-031-78113-1_11.

[44] W. Rahman, S. Roy, and A. Islam, "A comprehensive model for advanced road scene understanding: Yolo-cnn fusion for accurate road segmentation and object detection in varied conditions," Bachelor's thesis, Department of CSE, BRAC University, 2025.